



Catalogue 92-342F

# Échantillonnage et pondération

Rapports techniques du recensement de 1991



91

Recensement

Statistique Canada
Statistics Canada

Canada

## Pour nous joindre

### Des données sous plusieurs formes...

Statistique Canada diffuse les données sous formes diverses. Outre les publications, des totalisations habituelles et spéciales sont offertes. Les données sont disponibles sur disque compact, disquette, imprimé d'ordinateur, microfiche et microfilm, et bande magnétique. Des cartes et d'autres documents de référence géographiques sont disponibles pour certaines sortes de données. L'accès direct à des données agrégées est possible par le truchement de CANSIM, la base de données ordinolingue et le système d'extraction de Statistique Canada.

### Comment obtenir d'autres renseignements

Toutes les demandes de renseignements au sujet des produits et services doivent être adressées au centre de consultation de Statistique Canada de votre région :

| | | | |
|---|---|---|---|
| Halifax | (902 426-5331) | Regina | (306 780-5405) |
| Montréal | (514 283-5725) | Edmonton | (403 495-3027) |
| Ottawa | (613 951-8116) | Calgary | (403 292-6717) |
| Toronto | (416 973-6586) | Vancouver | (604 666-3691) |
| Winnipeg | (204 983-4020) | | |

Un service d'appel interurbain sans frais est offert, dans toutes les provinces et dans les territoires, **aux utilisateurs qui habitent à l'extérieur des zones de communication locale** des centres régionaux de consultation.

| | |
|---|---|
| Terre-Neuve et Labrador, Nouvelle-Écosse, Nouveau-Brunswick et Île-du-Prince-Édouard | 1-800-565-7192 |
| Québec | 1-800-361-2831 |
| Ontario | 1-800-263-1136 |
| Saskatchewan | 1-800-667-7164 |
| Manitoba | 1-800-661-7828 |
| Sud de l'Alberta | 1-800-882-5616 |
| Alberta et les Territoires du Nord-Ouest | 1-800-563-7828 |
| Colombie-Britannique et le territoire du Yukon | 1-800-663-1551 |
| **Appareil de télécommunications pour malentendants** | **1-800-363-7629** |
| **Numéro sans frais pour commander seulement (Canada et États-Unis)** | **1-800-267-6677** |

### Comment commander les publications

On peut se procurer les publications de Statistique Canada auprès des agents autorisés et des autres librairies locales, par l'entremise du centre régional de consultation de Statistique Canada le plus près, ou en écrivant à la Division du marketing, Ventes et Service, Statistique Canada, Ottawa K1A 0T6. Veuillez communiquer avec le centre régional de consultation le plus près pour de plus amples renseignements sur les produits et services électroniques. Voir le bon de commande à la fin du présent produit pour d'autres détails.

(1-613-951-7277)
Numéro du télécopieur (1-613-951-1548)
Toronto Carte de crédit seulement (1-416-973-8018)

### Normes de service au public

Afin de maintenir la qualité du service au public, Statistique Canada observe des normes établies en matière de produits et de services statistiques, de diffusion d'information statistique, de services à recouvrement des coûts et de services aux répondants. Pour obtenir une copie de ces normes de service, veuillez communiquer avec le Centre de consultation régional de Statistique Canada le plus près de chez vous.



Statistique Canada

# Rapports techniques du recensement de 1991

## Échantillonnage et pondération

## Série des produits de référence

Publication autorisée par le ministre responsable de Statistique Canada



Septembre 1994

Prix : Canada : 20 $
États-Unis : 24 $ US
Autres pays : 28 $ US

Catalogue 92-342F

ISBN 0-660-93600-3

Ottawa

This publication Is available in English upon request (92-342E)

---

**Note de reconnaissance**

*Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises et les administrations canadiennes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.*

Données de catalogage avant publication (Canada)

Vedette principale au titre :

Échantillonnage et pondération

(Série des produits de référence)
(Rapport technique du recensement de 1991)
Publié aussi en anglais sous le titre : Sampling and Weighting
ISBN 0-660-93600-3
CS92-342F

1. Canada -- Recensement, 1991 -- Méthodologie.
2. Canada -- Recensement, 1991 -- Évaluation.
I. Statistique Canada. II. Collection.

HA741.5.1991 S2514 1994 304.6'0723
C94-988075-2

**Pour citer ce document**

Statistique Canada. Échantillonnage et pondération.
Rapports techniques du recensement de 1991; Série des produits de référence.
Ottawa : Ministère de l'Industrie, des Sciences et de la Technologie, 1994.
Numéro 92-342F au catalogue.

Le papier utilisé dans le présent document répond aux exigences minimales de l'American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials. ANSI Z39.48-1984
♾

## Préface

Au fil des ans, le recensement du Canada est devenu la principale source de renseignements sur les Canadiens et leur mode de vie. Les décisions fondées sur cette source influencent la vie sociale et économique de tous les Canadiens.

À titre d'organisme chargé de produire ces renseignements, Statistique Canada doit informer les utilisateurs sur la qualité des données. Il incombe au Bureau d'exposer les concepts et la méthodologie ayant servi à la collecte et au traitement, ainsi que toute autre caractéristique des données susceptible d'avoir une incidence sur leur utilisation ou leur interprétation.

Afin de décrire la qualité des données du recensement de 1991, Statistique Canada a élaboré les publications suivantes : un **Dictionnaire** du recensement, qui présente de façon concise, sous forme de textes et de graphiques faciles à comprendre, de l'information sur les concepts du recensement; le **Recensement en bref**, qui fournit un aperçu des étapes du recensement; et une série de **Rapports techniques**, qui étudient en profondeur la qualité des données relatives à certaines caractéristiques, en l'occurrence le lieu de travail.

Les renseignements sur la qualité des données sont importants pour les utilisateurs. Ils leur permettent de déterminer si les données du recensement peuvent leur être utiles et d'évaluer les risques qu'il y a à fonder des conclusions ou des décisions sur ces données. Le recensement de 1991 a constitué une entreprise complexe et de grande envergure. Bien qu'on ait déployé des efforts considérables pour assurer le respect de normes élevées tout au long des opérations de collecte et de traitement, il est inévitable que les données résultantes soient entachées d'une certaine erreur.

Les renseignements sur la qualité des données revêtent également de l'importance pour Statistique Canada. Ils font partie intégrante de l'élaboration et du maintien de programmes statistiques pertinents et fiables.

La présente publication contribue largement à l'atteinte de ces objectifs. Elle a été réalisée par Mark Majkowski de la Section des opérations du recensement de la Division des méthodes d'enquêtes sociales. Les membres du personnel de deux divisions de Statistique Canada ont aussi contribué à la réalisation de cette publication, notamment ceux de la Division des méthodes d'enquêtes sociales et de la Division des opération du recensement.

En terminant, je tiens à remercier les millions de Canadiens et de Canadiennes qui ont accepté de remplir leur questionnaire du recensement le 4 juin 1991, de même que tous ceux et celles qui ont aidé Statistique Canada à planifier et à réaliser le recensement.

Ivan P. Fellegi

Chief Statistician of Canada

# Table des matières

# Liste des tableaux

# I. Introduction

De nos jours, l'échantillonnage constitue une technique largement utilisée dans maints domaines. Nous pouvons juger de la qualité d'un produit à l'aide d'un échantillon avant d'en faire l'achat; nous formons une opinion des gens à partir d'échantillons de leur comportement; nos impressions sur tel pays ou telle ville sont basées sur les courts voyages que nous y avons faits. Ce sont là des exemples d'échantillonnage où l'on tire des conclusions sur le «tout» à partir de connaissances portant sur une «partie».

D'une façon plus scientifique, l'échantillonnage est utilisé par les comptables, par exemple, lorsqu'ils effectuent la vérification des états financiers; dans l'industrie, pour vérifier la qualité d'articles produits en série, et par les initiateurs de sondages d'opinion et d'enquêtes pour recueillir des renseignements sur les opinions ou les caractéristiques d'une population donnée. On a généralement recours à l'échantillonnage pour réduire les coûts ou pour obtenir rapidement des données, ou les deux. Dans certains cas, le mesurage peut détruire le produit à l'étude (par exemple, l'évaluation de la durée de vie des ampoules électriques) et par conséquent, l'échantillonnage s'avère essentiel. Le désavantage de cette méthode est que les résultats obtenus à partir d'un échantillon peuvent ne pas être aussi précis que ceux recueillis auprès de la population totale. Toutefois, l'échantillonnage peut être avantageux quand la perte de précision (qui peut être minime si l'échantillon est suffisamment grand) est acceptable en fonction de l'utilisation prévue des résultats. De plus, la réduction de l'envergure d'une étude faite à partir d'un échantillon peut mener à la réduction des erreurs non dues à l'échantillonnage, ce qui compense en partie le manque de précision attribuable à l'échantillonnage.

Pour le recensement de la population de 1991, on a appliqué les méthodes d'échantillonnage de plusieurs façons. On y a eu recours au moment de tester le libellé des questions à l'étape de l'élaboration du questionnaire, ainsi que pour s'assurer que la qualité du travail des recenseurs dans le cadre de la collecte des questionnaires respectait certaines normes; on y a eu recours pour le contrôle qualitatif du codage des réponses au cours du dépouillement dans les bureaux régionaux; on y a eu recours pour estimer le sous-dénombrement et le surdénombrement qui se sont produits pour diverses raisons; on y a eu recours pour évaluer la qualité des données du recensement. Toutefois, la principale application de l'échantillonnage au recensement a été effectuée pendant le dénombrement sur le terrain où toutes les données, excepté les données de base, ont été recueillies auprès d'un échantillon de ménages. Le présent guide décrit cette dernière utilisation de l'échantillonnage et présente une évaluation de l'incidence de cette méthode sur la qualité des données du recensement.

Le chapitre II fait l'historique de l'application des méthodes d'échantillonnage aux recensements canadiens et décrit celles qui ont été utilisées au recensement de 1991. Le chapitre III décrit les méthodes de pondération des données-échantillon utilisées en vue d'obtenir des estimations à l'échelle de la population totale et explique les considérations d'ordre pratique et théorique qui régissent le choix de ces méthodes. On présente au chapitre IV les études effectuées dans le cadre du programme d'évaluation des méthodes d'échantillonnage et de pondération utilisées au recensement de 1991. Les chapitres V à VIII donnent les résultats de ces études.

# II. L'échantillonnage dans les recensements canadiens

Dans un recensement de la population, l'échantillonnage consiste à recueillir et à traiter certaines caractéristiques à partir d'un échantillon aléatoire de logements et de personnes dénombrés lors du recensement intégral. On obtient ensuite pour l'ensemble de la population les totaux pour ces caractéristiques en pondérant les données-échantillon. Les caractéristiques recueillies pour tous les logements ou toutes les personnes faisant partie du recensement sont appelées «caractéristiques de base» ou «caractéristiques 2A», tandis que les caractéristiques recueillies à partir d'un échantillon sont appelées «caractéristiques-échantillon» ou «caractéristiques 2B». Les désignations 2A et 2B représentent les questionnaires 2A et 2B, dont il est question à la section B ci-dessous.

## A. Historique de l'échantillonnage au recensement canadien[1]

L'échantillonnage a été utilisé pour la première fois au cours du recensement canadien de 1941. Un questionnaire sur le logement a été remis à tous les 10 logements de chaque sous-district de recensement. Les réponses aux 27 questions de ce questionnaire ont été intégrées à celles du questionnaire sur la population et les ménages correspondant à ces logements, ce qui a permis de faire le recoupement des caractéristiques-échantillon et des caractéristiques de base. Lors de ce même recensement, on a utilisé l'échantillonnage au cours de l'étape du dépouillement pour obtenir des estimations provisoires du revenu des salariés, de la répartition de la population en âge de travailler et de la composition des familles au Canada. Dans ce cas, l'échantillon était constitué du dixième de tous les secteurs de dénombrement du Canada et tous les questionnaires sur la population de ces secteurs ont été dépouillés à l'avance.

L'échantillonnage a également été utilisé pour le recensement du logement de 1951. Cette fois, on a choisi chaque cinquième logement (ceux dont le numéro d'identification se terminait par 2 ou par 7) et le questionnaire sur le logement comportait 24 questions. Au cours du recensement de 1961, on a prélevé un échantillon de 20 % des ménages privés et on a demandé aux membres de 15 ans et plus de ces ménages de remplir un questionnaire-échantillon sur la population comportant des questions sur la migration interne, la fécondité et le revenu. L'échantillonnage n'a pas été utilisé au cours des recensements de moindre envergure de 1956 et 1966.

Plusieurs innovations importantes ont été apportées aux méthodes de recensement en 1971. La première a été l'utilisation de l'autodénombrement pour la majorité de la population au lieu de la méthode traditionnelle de recensement par représentant. Ce changement découlait des résultats de plusieurs études effectuées au Canada et ailleurs (Fellegi, 1964, Hansen et autres, 1959) selon lesquels la présence d'un représentant avait un effet important sur la variance[2] de chiffres du recensement. On prévoyait que l'autodénombrement permettrait de réduire la variance des chiffres du recensement attribuable à la présence du représentant et, en même temps, donnerait au répondant le temps voulu pour remplir le questionnaire à son aise; par conséquent, on s'attendait à ce que les réponses soient plus exactes que par le passé.

Un autre changement survenu dans le recensement de 1971 a été la modification du contenu. Le nombre de sujets traités et de questions posées était plus élevé qu'aux recensements précédents. Une étude des coûts, du fardeau du répondant et des délais de production par rapport à la qualité des données dans le cadre d'un recensement par échantillonnage et par autodénombrement a présidé à la décision de recueillir toutes les données, sauf celles portant sur quelques caractéristiques de base, auprès d'un échantillon d'un tiers de la population au recensement de 1971. Sauf dans les régions les plus isolées du Canada, le tiers des ménages privés a reçu le questionnaire complet comportant toutes les questions du recensement, tandis que le reste des ménages privés a reçu le questionnaire abrégé qui ne contenait que les questions de base soit le nom, le lien de parenté avec le chef de ménage, le sexe, la date de naissance, l'état matrimonial, la langue maternelle, le type de logement, le mode d'occupation, le nombre de pièces, l'alimentation en eau, les toilettes ainsi que certaines questions en rapport avec la couverture du recensement. Tous les ménages vivant dans des régions éloignées prédésignées et dans des logements collectifs[3] ont

[1] Des renseignements détaillés sur les recensements particuliers sont contenus dans le **rapport administratif, la revue générale, le guide sommaire** ou le **recensement en bref** des recensements en question. Ces rapports sont mentionnés en bibliographie à la fin de ce guide.

[2] La «variance» d'une estimation est une mesure de sa précision. La variance est expliquée plus en détail au chapitre VIII.

[3] Un logement collectif est une institution ou un établissement commercial ou communautaire, par exemple les hôtels, les hôpitaux, les résidences de personnel et les camps de chantier.

reçu le questionnaire complet. Le rapport statistique «Sampling in the Census» (Bureau fédéral de la statistique, 1968) donne une description détaillée des projets d'application de cette méthode au recensement de 1971.

Le contenu du recensement de 1976 était beaucoup moindre que celui de 1971. En outre, le questionnaire de 1976 ne contenait pas les questions qui posent le plus de difficulté de collecte, par exemple les questions sur le revenu, ni celles pour lesquelles les frais de codage sont le plus élevés, par exemple celles sur la profession, l'activité économique et le lieu de travail. Par conséquent, les avantages de l'échantillonnage en termes de coûts et d'allégement du fardeau du répondant étaient moins évidents qu'au recensement de 1971. Néanmoins, après avoir évalué les économies éventuelles reliées à divers taux d'échantillonnage et étudié les conséquences sur le plan des relations publiques du retour au dénombrement intégral, il a été décidé d'appliquer en 1976 la même méthode d'échantillonnage qu'en 1971.

Au recensement de 1981, on a utilisé la plupart des mêmes méthodes qu'en 1971 et 1976, sauf que le taux d'échantillonnage est passé du tiers au cinquième des ménages. Les études réalisées à cette époque ont révélé que la diminution consécutive de la qualité des données (mesurée au moyen de la variance) serait acceptable et qu'elle ne serait pas suffisamment importante pour annuler les avantages découlant de la réduction des coûts et du fardeau du répondant, ainsi que la diminution des délais de production (voir Royce, 1983). Douze questions ont été posées à l'ensemble de la population et 34 questions de plus ont été posées aux membres de l'échantillon.

Le recensement de 1986 a été le premier recensement majeur réalisé au milieu d'une décennie. Il avait été décidé que seul un recensement majeur pourrait répondre aux besoins croissants en matière de données locales sur le marché du travail, besoins rendus plus pressants par l'importante récession (1981-1982) survenue depuis le recensement précédent. Toutefois, afin de maintenir les coûts d'élaboration au niveau le plus bas possible, on a décidé de faire très peu de changements. Par conséquent, les questions ainsi que les méthodes de collecte et de dépouillement des données du recensement de 1981 ont été reprises, sauf s'il y avait des raisons importantes de ne pas procéder ainsi. Des questions portant sur huit sujets posées au recensement de 1981 n'ont pas été reprises en 1986, tandis que trois nouvelles questions ont été ajoutées. Après la collecte des questionnaires du recensement de 1986, un échantillon de répondants a été prélevé pour la réalisation de l'enquête postcensitaire sur la santé et les limitations d'activité (ESLA). Cette enquête, réalisée pour la première fois en 1986, visait à brosser un tableau complet des conditions de vie des personnes ayant une incapacité.

En 1991, le recensement de la population englobait à la fois les résidents permanents et non permanents[4] du Canada. À l'exception du recensement de 1941, seuls les résidents permanents du Canada étaient inclus dans le recensement avant 1991. Afin qu'on puisse savoir qui étaient les résidents non permanents, une nouvelle question a dû être élaborée et incluse dans le recensement de 1991. En tout, 12 nouvelles questions ont été ajoutées au recensement de 1991, tandis que les questions relatives à quatre sujets du recensement de 1986 n'ont pas été posées en 1991. Des 12 nouvelles questions, sept étaient posées pour la toute première fois, tandis que cinq étaient reprises de recensements antérieurs. Une fois terminée la collecte des questionnaires du recensement de 1991, deux enquêtes postcensitaires ont été effectuées en 1991. Il s'agissait de l'ESLA (aussi réalisée en 1986) et de l'Enquête auprès des peuples autochtones (EAPA). Cette dernière enquête, effectuée pour la première fois en 1991, visait à recueillir des données sur les populations autochtones vivant dans les réserves et hors des réserves. On a également noté, à l'occasion du recensement de 1991, une augmentation importante de l'automatisation du traitement des données et des véhicules par lesquels les produits et les services sont élaborés et livrés au client.

## B. Plan d'échantillonnage utilisé lors du recensement de 1991

Le jour du recensement en 1991, une multitude de renseignements ont été recueillis auprès de chaque Canadien. La majeure partie de ces informations ont été obtenues à partir d'un échantillon. Dans chaque secteur d'autodénombrement, quatre logements privés occupés sur cinq ont reçu un questionnaire abrégé (questionnaire 2A) contenant neuf questions de base sur l'âge, le sexe, l'état matrimonial, l'union libre, la langue maternelle, le lien de parenté avec la personne repère du ménage (Personne 1), le type de logement et le mode

---

4 Les résidents non permanents sont des personnes qui sont titulaires de permis de séjour pour étudiant ou employé ou d'autorisations ministérielles, ou encore qui ont demandé le statut de réfugié.

d'occupation. Un ménage sur cinq a reçu le questionnaire complet (questionnaire 2B) contenant les neuf questions de base, ainsi que 44 autres questions de nature socio-économique ou liées au logement.

Tous les logements se trouvant dans les secteurs dénombrés par un recenseur (généralement les régions éloignées ou les réserves indiennes) ont reçu le questionnaire 2D. Le questionnaire 2D avait le même contenu que le questionnaire 2B (sauf pour la question portant sur le mode d'occupation du logement), mais il était conçu pour être utilisé dans le cadre d'une interview en personne.

Un questionnaire 2B a aussi été créé pour tous les logements collectifs. Toutefois, les résidents[5] des logements collectifs institutionnels n'ont pas eu à répondre au questionnaire-échantillon, et seule l'information de base a été recueillie dans leur cas. Les membres du personnel vivant dans ces logements collectifs institutionnels, les personnes vivant dans des logements collectifs non institutionnels (y compris le personnel résident), ainsi que les Canadiens postés à l'extérieur du Canada (généralement le personnel des ambassades ou des Forces armées) ont dû fournir l'information du questionnaire complet, sauf pour les questions portant sur le logement. Toutefois, les Canadiens postés à l'extérieur du Canada devaient répondre à des questions sur le lieu habituel de résidence de la personne au Canada. Les renseignements relatifs aux logements privés inoccupés ont été consignés sur le questionnaire 2A.

D'après les procédures de livraison, le recenseur devait planifier un itinéraire couvrant tous les logements de son secteur de dénombrement (SD) puis se rendre à chaque logement et y laisser un questionnaire. Le choix de l'échantillon, c'est-à-dire la décision quant au genre de questionnaire à laisser dans chaque logement occupé, a été fait à partir du Registre des visites (RV) dans lequel le recenseur inscrivait tous les logements de son secteur. Chaque cinquième ligne du registre était ombrée, indiquant qu'un questionnaire 2B devait être laissé. Un point de départ aléatoire était déterminé à partir de la première ligne du RV, et le recenseur sautait zéro, une, deux, trois ou quatre lignes selon que le cinquième, quatrième, troisième, deuxième ou premier logement du SD devait être le premier à recevoir le questionnaire complet. Par la suite, chaque logement inscrit sur une ligne ombrée recevait automatiquement le questionnaire complet. Ces procédures étaient expliquées dans le manuel du recenseur (Formule 41) et une attention particulière y était portée pendant la formation du recenseur afin de réduire au minimum le risque d'erreur dans la sélection de l'échantillon. Des vérifications du travail des recenseurs, à des fins de contrôle qualitatif, ont été effectuées par le commissaire au recensement.

Dans la terminologie de l'échantillonnage, l'échantillon peut être décrit comme un échantillon stratifié avec sélection systématique de logements privés occupés, choisi en fonction d'un taux d'échantillonnage constant de 1 sur 5 dans toutes les strates (SD). En tant qu'échantillon de personnes, il peut être considéré comme un échantillon stratifié avec sélection systématique de grappes, un logement constituant une grappe. Stuart (1976) et Cochran (1977) donnent une description détaillée des concepts et de la terminologie se rapportant à l'échantillonnage.

## C. Traitement des données de l'échantillon du recensement

Dès que le recenseur était en possession du questionnaire rempli (2A, 2B ou 2D) de chaque logement de son secteur et que son travail avait été approuvé, les questionnaires étaient envoyés à l'un des sept sites de dépouillement régionaux aux fins du traitement manuel. À ces endroits, les questionnaires étaient enregistrés, comptés et préparés pour l'introduction sur clavier des données. La préparation incluait des vérifications de cohérence entre les questionnaires et le Registre des visites, ainsi que des vérifications de lisibilité visant à assurer que les documents se prêtaient à une entrée des données sur ordinateur. En outre, les réponses écrites à cinq questions ont été converties en codes numériques permettant l'entrée directe des données. La transcription des données des formules 4A (créées pour les ménages manquants ou refusant de répondre) vers une formule 2A ou 2B, ainsi que la transcription de l'information du questionnaire complet recueillie auprès des personnes postées à l'extérieur du Canada ou vivant dans des logements collectifs vers une formule 2B, ont également été faites à ces sites. Les données complètes pour chaque SD ont été saisies et stockées sur bandes magnétiques. Les questionnaires et les bandes magnétiques ont ensuite été envoyés au bureau central, à Ottawa.

---

5 Détenus dans les établissements de correction et établissements pénitentiaires ou les prisons; patients dans les hôpitaux et les patients dans les hôpitaux pour personnes souffrant de maladies chroniques, ou les hôpitaux psychiatriques; les résidents des foyers pour personnes âgées; enfants dans les foyers collectifs pour enfants (orphelinats) ou les foyers pour jeunes contrevenants.

Au bureau central, des contrôles de structure automatisés étaient effectués au niveau des secteurs de dénombrement, des ménages et des personnes, et certaines incohérences, par exemple des divergences entre les nombres de personnes ou entre les nombres de ménages selon les niveaux géographiques, étaient résolues manuellement. Un processus de codage automatique convertissait en codes numériques les réponses écrites à plusieurs questions. Pour la première fois, les correspondances étaient établies automatiquement entre les réponses écrites saisies reçues du bureau central et une structure de classification/fichier de référence automatisée. Cette structure comprenait une série de mots ou d'expressions, ainsi que des codes numériques connexes pour chacune de ces variables. À la fin du traitement au bureau central, les ménages 2B n'ayant pas répondu à toutes les questions propres au questionnaire 2B ont été convertis en ménages 2A. La taille de l'échantillon s'en est trouvée réduite, ce qui a eu pour effet d'accroître les coefficients de pondération des ménages 2B restants. On a considéré qu'on obtiendrait de la sorte des estimations de meilleure qualité que si toutes les réponses 2B pour ces ménages avaient été imputées. Une fois terminées toutes les mises à jour des données pour un SD, les données étaient restructurées et transférées à la section Contrôle et Imputation.

Les données ont été chargées dans 10 bases de données de Contrôle et Imputation, soit cinq bases pour les données 2A (100 %) et cinq bases pour les donnés 2B (20 %). Dans chaque cas, quatre des cinq bases de données correspondaient aux quatre régions géographiques du Canada (Est, Québec, Ontario et Ouest), tandis que la cinquième englobait les Canadiens postés à l'extérieur du Canada (groupe appelé 2C). Les bases de données 2A contenaient les caractéristiques démographiques de base pour l'ensemble de la population et les bases de données 2B contenaient les données-échantillon (20 %). Les données ont été soumises à une série de modules de traitement spécialisés, où tous les problèmes liés à des données invalides, incohérentes ou manquantes ont été résolus. Les bases de données 2A ont été traitées en premier et ont donné lieu à la création de la base de données d'extraction 2A définitive.

Une fois la dernière main mise aux données intégrales, les données-échantillon (20 %) ont été traitées. Une base de données d'extraction 2B définitive a été créée. Cette base de données contenait à la fois les données intégrales et les données-échantillon uniquement pour les ménages et les personnes de l'échantillon. Les coefficients de pondération calculés à l'aide des données intégrales (de la façon décrite au chapitre III) ont été chargés dans cette base de données.

# III. Estimations basées sur l'échantillon du recensement

Tout échantillonnage nécessite l'application d'une méthode d'estimation pour permettre une inférence sur la population totale à partir des données-échantillon. La méthode d'estimation est généralement choisie en fonction d'exigences d'ordre pratique et théorique. Sur le plan pratique, cette méthode doit être compatible avec le système de dépouillement auquel elle s'intègre, tandis que sur le plan théorique, elle doit minimiser l'erreur d'échantillonnage des estimations produites. Les deux sections qui suivent décrivent les considérations pratiques et théoriques qui régissent le choix des méthodes d'estimation à partir de l'échantillon du recensement.

## A. Considérations d'ordre pratique

Une méthode d'estimation peut être décrite mathématiquement à l'aide d'une formule algébrique qui indique comment la valeur de l'estimateur de la population est calculée en fonction des valeurs observées dans l'échantillon. Quand il s'agit de petites enquêtes qui ne portent que sur une ou deux caractéristiques, ou, lorsque la formule d'estimation est très simple, il est possible de calculer les estimations-échantillon en appliquant cette formule aux données-échantillon pour chaque estimation requise. Toutefois, dans le cas d'une enquête ou d'un recensement portant sur une gamme étendue de caractéristiques, ou lorsque la formule d'estimation est assez complexe, il n'est pas possible d'appliquer une formule séparément pour chaque estimation désirée. Dans le cas du recensement, par exemple, chaque cellule de chaque totalisation, fondée sur des données-échantillon recueillies à chaque niveau géographique, représente une estimation-échantillon qui nécessiterait, selon cette approche, une application distincte de la formule d'estimation. Par ailleurs, le fait de calculer séparément chaque estimation ne mènerait pas nécessairement à la cohérence des diverses estimations effectuées à partir d'un même échantillon.

Par conséquent, la méthode utilisée pour le recensement et pour de nombreuses enquêtes-échantillon consiste à décomposer le processus d'estimation en deux étapes : a) le calcul des coefficients de pondération, c'est-à-dire la pondération elle-même et b) la sommation des coefficients de pondération afin de produire des estimations des chiffres de population. Les seules difficultés mathématiques, s'il y en a, se rapportent à l'étape a) qui n'est effectuée qu'une fois, tandis que l'étape b) se limite à un simple processus de sommation des coefficients de pondération au moment où une totalisation est extraite. De plus, comme le coefficient de pondération relié à une unité de l'échantillon est le même quelle que soit la totalisation effectuée, la cohérence entre les diverses estimations basées sur les données-échantillon est assurée.

## B. Considérations d'ordre théorique

Selon la théorie de l'échantillonnage, il est possible de déterminer, à partir d'un plan d'échantillonnage particulier et d'une procédure d'estimation donnée, les chances qu'un certain intervalle contienne la valeur (inconnue) que l'on veut estimer pour la population entière. Le principal objet de l'estimation est de réduire au maximum l'étendue de ces intervalles de façon que les hypothèses sur les valeurs inconnues de la population soient aussi précises que possible. La mesure courante de précision utilisée pour comparer les procédures d'estimation est appelée l'erreur-type. Pourvu que certaines conditions relativement peu contraignantes soient respectées, un intervalle de plus ou moins deux erreurs-types par rapport à la valeur estimée renfermera la valeur de la population pour environ 95 % de tous les échantillons possibles.

Outre la réduction au minimum de l'erreur-type, le deuxième objectif relatif au choix d'une méthode d'estimation de l'échantillon du recensement est de garantir autant que possible que les estimations faites à partir de l'échantillon des caractéristiques de base (c.-à-d. 2A) soient cohérentes avec les valeurs connues de la population correspondante. Heureusement, ces deux objectifs sont généralement complémentaires, c'est-à-dire que l'erreur d'échantillonnage est habituellement réduite lorsqu'on s'assure que les estimations-échantillon de certaines caractéristiques de base sont compatibles avec les chiffres correspondants de la population. Notons cependant, qu'il peut arriver que le fait de forcer les estimations-échantillon de caractéristiques de base à être cohérentes avec les chiffres correspondants de la population pour de très petits sous-groupes puisse avoir des conséquences défavorables sur l'erreur-type des estimations des caractéristiques-échantillon proprement dites.

Lorsqu'on n'a aucune information sur la population échantillonnée autre que celle qui a été recueillie pour les unités d'échantillonnage, la méthode d'estimation se limite à pondérer les unités d'échantillonnage en proportion inverse de leur probabilité de sélection; par exemple, si toutes les unités avaient une chance sur 5 d'être choisies, toutes les unités recevraient un coefficient de pondération de 5. En pratique, toutefois, on a généralement une certaine connaissance de la population observée, par exemple, sa taille globale et, peut-être, sa subdivision en fonction d'une variable donnée (disons, par province). Ces données peuvent être utilisées pour améliorer la formule d'estimation de façon à produire des estimations qui se rapprochent davantage de la valeur inconnue de la population. Dans le cas de l'échantillon du recensement, on dispose d'un grand nombre de données très détaillées sur la population échantillonnée, soit les caractéristiques de base obtenues pour tous les ménages, à tous les niveaux géographiques. D'une part, on peut se servir de cette information pour améliorer les estimations faites à partir de l'échantillon du recensement mais d'autre part, cette multitude de données peut également occasionner des ennuis compte tenu du fait qu'il est impossible d'obtenir des estimations-échantillon des caractéristiques de base qui soient cohérentes avec toutes les données sur la population à tous les niveaux géographiques. Les écarts entre les estimations-échantillon faites à partir de l'échantillon et les valeurs de la population apparaissent quand on produit un recoupement d'une variable de base et d'une variable-échantillon. La totalisation doit être fondée sur les données-échantillon; de ce fait, les totaux marginaux des variables de base sont des estimations-échantillon que l'on peut comparer avec les chiffres correspondants de la population d'une autre totalisation basée sur les données intégrales. Il se peut que cette comparaison ne donne pas une concordance exacte.

## C. Élaboration d'une méthode d'estimation pour l'échantillon du recensement

Étant donné qu'il faut attribuer un coefficient de pondération à chaque unité (personne, famille ou ménage) de l'échantillon, le plus simple serait de choisir le coefficient 5, puisqu'il s'agit d'un échantillon de 1 sur 5. Une telle méthode serait simple et sans biais[6], et si l'on n'avait aucune autre donnée que celles de l'échantillon, cela pourrait être la meilleure méthode. Or, bien qu'on sache que l'échantillon contiendra presque exactement le cinquième de tous les ménages (à l'exclusion des ménages collectifs et de ceux qui se trouvent dans les secteurs de recensement par interview), on ne peut pas être certain qu'il contiendra exactement un cinquième de toutes les personnes, ou un cinquième de chaque genre de ménage, ou un cinquième de toutes les femmes de 25 à 34 ans, etc. Par conséquent, une telle méthode n'assurerait pas la cohérence, même pour les sous-groupes les plus importants de la population. Dans ce dernier cas, les fractions seraient très proches d'un cinquième, mais elles pourraient s'en écarter de façon sensible dans le cas des petits sous-groupes. Une autre méthode simple serait de prendre certains sous-groupes importants, par exemple les groupes d'âge-sexe par province, puis, pour chacun, compter le nombre d'éléments de la population appartenant au sous-groupe (N) et le nombre d'unités de l'échantillon appartenant au sous-groupe (n), et attribuer à chaque unité d'échantillonnage du sous-groupe un coefficient de pondération égal à N/n.

Par exemple, si l'on recensait 5 000 hommes de 20 à 24 ans à l'Île-du-Prince-Édouard, et si 1 020 de ces hommes faisaient partie des ménages inclus dans l'échantillon, un coefficient de pondération de 5 000/1 020 = 4,90 serait attribué à chaque homme de 20 à 24 ans dans l'échantillon de l'Île-du-Prince-Édouard. De cette façon, chaque fois que les groupes d'âge-sexe quinquennaux seraient recoupés avec une caractéristique-échantillon pour l'Île-du-Prince-Édouard, le total marginal pour le groupe d'âge-sexe d'hommes de 20 à 24 ans concorderait avec le total de la population, soit 5 000. Ce genre de méthode d'estimation est appelée «estimation par quotient». Il est à noter que, dans cet exemple particulier, un coefficient de pondération de 5 donnerait une estimation-échantillon de 5 100 (1 020 x 5). La méthode d'estimation qui a été utilisée dans le recensement de 1986 était une généralisation de l'estimation par quotient appelée méthode itérative par le quotient (MIQ). Pour plus de détails sur la MIQ, voir le Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986: Échantillonnage et pondération, ainsi que Brackstone et Rao (1979).

Pour le recensement de 1991, il a été décidé de recourir à une autre méthode d'estimation, appelée procédure d'estimation généralisée par les moindres carrés (PEGMC) en deux étapes. Ce choix visait à améliorer la concordance entre les totaux de la population et les estimations correspondantes au niveau des SD, par rapport à ce que permettait la MIQ. Les erreurs-types des estimations produites par la PEGMC pour les petites régions géographiques étaient également réduites. En outre, la PEGMC permettait de déterminer, pour chaque ménage de

6 «Sans biais» veut dire que la moyenne des estimations de l'ensemble de tous les échantillons possibles serait égale à la valeur vraie de la population.

l'échantillon, un coefficient de pondération unique pouvant servir à produire des estimations tant pour les caractéristiques des personnes que pour les caractéristiques des ménages. Avec la MIQ, il fallait utiliser des coefficients de pondération différents pour produire des estimations des caractéristiques des ménages et des personnes, ce qui entraînait parfois des incohérences. Il y avait aussi parfois des incohérences résultant du fait que le processus itératif de la MIQ pour le calcul des coefficients de pondération ne convergeait pas toujours (voir Daoust et Bankier, 1989).

Avec la PEGMC (qui correspond, on peut le montrer, à un estimateur de régression), les coefficients de pondération d'environ 5 ont été ajustés aussi peu que possible pour chaque ménage, en même temps qu'on s'assurait qu'il y avait une concordance parfaite entre les estimations et les chiffres de population pour autant de caractéristiques de base que possible. Ces soi-disantes «contraintes» sont énoncées à l'annexe A. Il fallait que cette concordance parfaite soit réalisée au niveau des régions de pondération (RP). Chaque RP contenait, en moyenne, sept SD échantillonnés. Pour plus de détails sur les RP, voir le chapitre VI, section A, du présent rapport.

## D. Procédure d'estimation généralisée par les moindres carrés en deux étapes

Les calculs des coefficients de pondération sont faits indépendamment dans chaque RP. Certaines des contraintes indiquées à l'annexe A (tant au niveau des SD qu'à celui des RP) doivent être retranchées pour chaque RP, de sorte que la concordance entre les chiffres de population et les estimations ne peut être garantie pour toutes les contraintes. Les contraintes sont initialement retranchées au niveau des RP pour les raisons suivantes :

- elles s'appliquent à moins de 60 ménages (on les appelles des contraintes faibles);
- elles sont redondantes (on les appelle des contraintes linéairement dépendantes [LD]);
- elles sont quasi redondantes (on les appelle des contraintes quasi linéairement dépendantes [QLD]).

Par exemple, puisque le nombre total de femmes plus le nombre total d'hommes égale la population totale, le nombre total de femmes peut être retranché à titre de contrainte redondante ou «linéairement dépendante», car il suffit que deux contraintes sur trois, n'importe lesquelles, soient satisfaites pour assurer que la troisième le sera également. Pour illustrer ce qu'on entend par contrainte quasi redondante, prenons le cas des contraintes représentant les personnes dont l'état matrimonial est «personne séparée» et les soutiens du ménage dont l'état matrimonial est «personne séparée». Si la plupart, mais non la totalité, des personnes séparées sont des soutiens du ménage, alors les deux contraintes sont presque équivalentes, et l'on peut en considérer une comme QLD. Les contraintes LD ont été retranchées pour accroître l'efficacité de calcul de l'algorithme de pondération. Les contraintes faibles et les contraintes QLD ont été retranchées parce que si on ne l'avait pas fait, les estimations auraient pu devenir instables et présenter des erreurs-types élevées.

Une fois retranchées les contraintes faibles, LD et QLD au niveau des RP, le calcul des coefficients de pondération PEGMC a lieu en deux étapes. À la première étape, les coefficients de pondération initiaux, égaux à l'inverse de la fraction d'échantillonnage des ménages dans les SD, sont ajustés individuellement pour chaque SD. Certaines contraintes non retranchées au niveau des RP peuvent être retranchées au niveau des SD, parce qu'elles sont faibles ou linéairement dépendantes. Les contraintes qui demeurent pour le SD sont triées selon le nombre de ménages auxquels elles s'appliquent au niveau du SD. Elles sont ensuite divisées en deux groupes, soit celles de rang pair et celles de rang impair. Les coefficients de pondération PEGMC sont calculés au niveau du SD pour chaque groupe de contraintes. Parfois, la méthode d'estimation donne des coefficients de pondération très faibles (inférieurs à 1) ou très élevés (supérieurs à 25) pour procurer la concordance nécessaire à l'égard de certaines contraintes. Ces coefficients, appelés «valeurs aberrantes», ne sont pas désirables. Par conséquent, quand cela se produit, les contraintes responsables sont repérées et retranchées, et les coefficients sont calculés de nouveau. Enfin, on fait la moyenne des coefficients obtenus pour les deux groupes de contraintes pour chaque ménage de l'échantillon, ce qui donne les coefficients de pondération de la première étape pour chaque SD.

Les coefficients de pondération produits à la première étape sont utilisés comme coefficients initiaux de la deuxième étape, qui consiste à les ajuster de façon à obtenir une concordance entre les estimations-échantillon et les chiffres de population au niveau des RP. Toutes les contraintes non désignées comme faibles, LD ou QLD au niveau de la

RP sont utilisées. Ici encore, si des valeurs aberrantes sont produites, les contraintes responsables sont repérées et retranchées, et les coefficients finals sont calculés de nouveau. Même si la deuxième étape détruit quelque peu la concordance obtenue pour les estimations au niveau des SD à la première étape, les estimations finales au niveau des SD sont tout de même plus proches des chiffres de population que si la première étape avait été omise. En outre, les contraintes exigeant une concordance exacte pour le nombre total de ménages et la population totale au niveau des SD (voir l'annexe A) sont appliquées dans l'ajustement de pondération de la deuxième étape, à moins qu'elles soient retranchées parce qu'elles sont faibles, LD ou QLD, ou encore parce qu'elles entraînent des coefficients de pondération aberrants. Pour une explication plus détaillée du calcul des coefficients de pondération, voir Bankier, Rathwell et Majkowski (1992).

Les SD dans lesquels des questionnaires 2A et des questionnaires 2B ont été distribués aux logements privés occupés sont appelés SD échantillonnés. Les coefficients de pondération PEGMC ont été calculés seulement pour les ménages 2B de logements privés occupés dans les SD échantillonnés. Les ménages de logements privés occupés qui ont reçu un questionnaire 2A dans les SD échantillonnés se sont vu attribuer le coefficient de pondération 0. Tous les ménages de logements privés occupés dans les SD non échantillonnés ont reçu le coefficient 1, de même que tous les ménages habitant des logements collectifs, peu importe le type de SD auquel ils appartenaient.

# IV. Programme d'évaluation de l'échantillonnage et de la pondération

Le programme d'évaluation de l'échantillonnage et de la pondération avait pour objet de déterminer l'effet de l'échantillonnage et de la pondération sur la qualité des données-échantillon du recensement. À cette fin, cinq études ont été réalisées afin de mesurer la qualité des données-échantillon et des estimations-échantillon du recensement ainsi que de fournir des renseignements utiles à la planification des futurs recensements. Les études ont porté sur les sujets suivants :

a) examen du biais d'échantillonnage;

b) évaluation de la formation des régions de pondération;

c) évaluation des méthodes de pondération;

d) examen de la cohérence des estimations-échantillon et des chiffres de population;

e) étude visant à évaluer la variance d'échantillonnage pour diverses caractéristiques-échantillon.

La suite du présent chapitre décrit brièvement chacune de ces études. Les chapitres V à VIII en présentent les résultats.

## A. Examen du biais d'échantillonnage

Dans toute enquête, un biais entachant les réponses peut provenir de diverses sources. Le but de cette étude était de déterminer si les réponses aux questions de base du questionnaire 2B étaient biaisées de quelque façon que ce soit et de cerner, si possible, les causes de tout biais observé.

## B. Évaluation de la formation des régions de pondération

Cette étude visait à mesurer le degré de conformité des RP aux critères de formation établis. Toutes les RP au Canada ont fait l'objet d'une analyse en vue de déterminer dans quelle mesure elles respectaient les contraintes de taille et les limites des divers types d'unités géographiques.

## C. Évaluation des méthodes de pondération

Le but de cette étude était d'évaluer l'efficacité de la PEGMC. À cette fin, on a examiné le niveau de concordance entre les estimations-échantillon et les chiffres de population à l'égard des diverses contraintes, pour l'ensemble des RP au Canada. Le nombre et le type des contraintes retranchées au niveau des RP, ainsi que les raisons pour lesquelles elles avaient été retranchées, ont été étudiés pour expliquer les incohérences observées. En outre, on a examiné la distribution des coefficients de pondération PEGMC, ainsi que les écarts entre les résultats de 1991 et ceux de 1986.

## D. Examen de la cohérence des estimations-échantillon et des chiffres de population

Dans cette étude, on a examiné le niveau de concordance, ou de cohérence, entre les estimations-échantillon et les chiffres de population pour un large éventail de caractéristiques de base, et non seulement pour celles utilisées comme contraintes dans la PEGMC. On a étudié cette concordance pour diverses unités géographiques autres que les RP. On a également comparé la cohérence obtenue en 1991 à celle obtenue en 1986 pour ces caractéristiques.

## E. Variance d'échantillonnage

La «variance» d'une estimation est une mesure de sa précision. Il est relativement peu coûteux de calculer des estimations de la variance pour des estimateurs lorsqu'on utilise des coefficients de pondération égaux à 5 et qu'on

suppose un échantillonnage aléatoire simple. Toutefois, il est beaucoup plus coûteux d'estimer la variance des estimateurs du recensement pour le plan d'échantillonnage et les techniques d'estimation utilisés. Cette étude indique comment des «facteurs d'ajustement» ont été calculés pour le recensement de 1986, sous la forme du rapport entre les estimations des erreurs-types (les racines carrées de la variance) des estimations du recensement et les estimations simples des erreurs-types. On peut alors obtenir une estimation de l'erreur-type d'une estimation du recensement pour n'importe quelle caractéristique dans toute unité géographique en multipliant l'estimation simple de l'erreur-type par le facteur d'ajustement approprié. On explique ensuite pourquoi ces estimations de l'erreur-type ne sont pas nécessairement exactes, en raison du biais introduit dans le processus par l'échantillonnage, le traitement des données et la méthode d'estimation.

# V. Biais d'échantillonnage

## A. Introduction

Les estimations établies à partir d'une enquête-échantillon sont sujettes à des erreurs d'échantillonnage. Un type d'erreur d'échantillonnage est causé par la variabilité des caractéristiques de la population. Cette variabilité signifie que des échantillons différents produiront des estimations différentes, aucune d'elles n'égalant nécessairement la vraie valeur de la population. Toutefois, les estimations égaleront la vraie valeur de la population en moyenne pourvu que l'échantillon ne comporte pas de biais donnant lieu à une tendance à la surestimation ou à la sous-estimation. Malheureusement, il est souvent difficile d'éliminer le biais complètement. Dans le cas du recensement de la population, les réponses peuvent être entachées d'un biais causé, entre autres, par les erreurs de couverture, le biais de non-réponse, le biais de réponse (p. ex., des recensés qui répondent différemment sur la formule 2B que sur la formule 2A), les erreurs commises par les recenseurs (p. ex., des recenseurs ne choisissant pas l'échantillon selon les exigences établies), les erreurs de traitement, et ainsi de suite.

L'examen du biais d'échantillonnage a été effectué pour déterminer si les réponses aux questions de base de la formule 2B comportaient un biais. Les estimations-échantillon de 53 caractéristiques de base (l'annexe B explique le rapport entre ces caractéristiques et les contraintes indiquées à l'annexe A) fondées sur des données imputées ont été comparées aux chiffres de population pour les 284 divisions de recensement (DR) échantillonnées au Canada. On a obtenu ces estimations-échantillon en multipliant les chiffres de l'échantillon au niveau du SD par le coefficient de pondération simple égal à la fraction de sondage inverse des ménages du SD (environ 5) et en faisant ensuite la somme au niveau de la DR[7]. On a constaté que la différence moyenne entre les estimations-échantillon et les chiffres de population calculée sur l'ensemble des DR était statistiquement significative (au niveau de 5 %)[8] pour la plupart des caractéristiques (c.-à-d. que les différences ne peuvent être expliquées par la variabilité d'échantillonnage). La mesure a été faite au moyen de la statistique

$$Z^{(0)} = \frac{\hat{X}^{(0)} - X}{\sqrt{V(\hat{X}^{(0)})}} \qquad (1)$$

où $\hat{X}^{(0)}$ est une estimation fondée sur les coefficients de pondération simples du chiffre de population 2A connu $X$ et $V(\hat{X}^{(0)})$ est la variance d'échantillonnage de l'estimateur $\hat{X}^{(0)}$. Les valeurs $Z^{(0)}$, pour les 284 DR, devraient suivre à peu près une distribution normale de moyenne 0 et de variance 1 si un échantillon aléatoire simple de ménages a été prélevé sans biais dans chaque SD et n'a pas été perturbé par le traitement (voir l'annexe C pour plus de détails).

## B. Principaux résultats

Le tableau 1 indique les différences (chiffres absolus et pourcentages) entre les estimations-échantillon et les chiffres de population à l'échelle nationale pour l'ensemble des 53 caractéristiques 2A. Dans la plupart de cas, le biais était inférieur à 1 %. Toutefois, il y a 43 caractéristiques marquées d'un astérisque au tableau 1 pour lesquelles les différences se sont révélées statistiquement significatives au niveau de 5 % d'après la statistique $Z^{(0)}$, dont les valeurs sont données au tableau 2. (Il convient de signaler que les nombres et les pourcentages du tableau 1 du Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986 : Échantillonnage et pondération étaient erronés et auraient dû être multipliés par un facteur de 2,6.)

---

7 Ces estimations simples ont été utilisées au lieu des estimations PEGMC, parce que la PEGMC tend à masquer le biais d'échantillonnage en forçant les estimations de caractéristiques de base à égaler les chiffres de population.

8 Cela signifie que la possibilité de relever de si grandes différences en l'absence d'un biais était d'au plus 5 %.

**Tableau 1. Estimation-échantillon (coefficients simples) moins chiffre de population à l'échelle du Canada (SD échantillonnés seulement) et pourcentage des DR dans lesquelles la caractéristique était surreprésentée**

| Caractéristiques étudiées | Estimation-échantillon moins chiffre de population | | | Pourcentage des DR avec surreprésentation |
|---|---|---|---|---|
| **Personnes** | Valeur | | Pourcentage | |
| Hommes | 3 275 | | (+0,03%) | 63 |
| Femmes | 64 216 | | (+0,48%) | 78 |
| Population totale | 67 491 | | (+0,26%) | 76 |
| Âge, 0-4 | 16 950 | * | (+0,92%) | 63 |
| Âge, 5-9 | 21 031 | * | (+1,14%) | 68 |
| Âge, 10-14 | 21 376 | * | (+1,17%) | 64 |
| Âge, 15-19 | 8 115 | * | (+0,45%) | 59 |
| Âge, 20-24 | -16 841 | * | (-0,89%) | 43 |
| Âge, 25-29 | -16 727 | * | (-0,73%) | 46 |
| Âge, 30-34 | 170 | * | (+0,01%) | 60 |
| Âge, 35-39 | 3 000 | * | (+0,13%) | 62 |
| Âge, 40-44 | 7 938 | * | (+0,39%) | 61 |
| Âge, 45-49 | 10 017 | * | (+0,63%) | 62 |
| Âge, 50-54 | 5 339 | * | (+0,41%) | 58 |
| Âge, 55-59 | 3 034 | * | (+0,26%) | 51 |
| Âge, 60-64 | 4 191 | * | (+0,37%) | 51 |
| Âge, 65-74 | 7 063 | * | (+0,39%) | 49 |
| Âge, > 74 | -7 165 | * | (-0,68%) | 35 |
| Célibataires | 12 | * | (+0,00%) | 56 |
| Personnes mariées | 95 348 | * | (+0,83%) | 85 |
| Veufs et veuves | -5 073 | * | (-0,42%) | 37 |
| Personnes divorcées | -15 198 | * | (-1,22%) | 38 |
| Personnes séparées | -7 598 | * | (-1,31%) | 39 |
| **Familles de recensement** | | | | |
| Nombre total de familles | 52 069 | * | (+0,72%) | 88 |
| Familles époux-épouse | 54 989 | * | (+0,87%) | 89 |
| Familles monoparentales | -2 921 | * | (-0,31%) | 45 |
| Enfants dans les familles | 72 463 | * | (+0,84%) | 75 |
| Personnes dans les familles | 179 522 | * | (+0,81%) | 85 |
| Personnes hors famille | -112 031 | * | (-2,75%) | 9 |
| **Ménages et logements** | | | | |
| Logements possédés | 46 713 | * | (+0,75%) | 81 |
| Logements loués | -46 713 | * | (-1,28%) | 19 |
| Maisons individuelles | 27 243 | * | (+0,49%) | 77 |
| App. dans un immeuble ≥ 5 étages | -931 | | (-0,10%) | 40 |
| Logements mobiles | -796 | | (-0,46%) | 60 |
| Autres logements | -25 516 | * | (-0,80%) | 26 |
| Nombre total de ménages | 0 | | (+0,00%) | 0 |
| Ménages d'une personne | -37 392 | * | (-1,66%) | 15 |
| Ménages de 2 personnes | 12 541 | * | (+0,40%) | 58 |
| Ménages de 3 personnes | 8 606 | * | (+0,50%) | 59 |
| Ménages de 4 personnes | 15 320 | * | (+0,88%) | 69 |
| Ménages de 5 personnes | 4 857 | * | (+0,68%) | 58 |
| Ménages de 6 personnes et plus | -3 932 | | (-1,25%) | 43 |
| Ménages non familiaux | -56 518 | * | (-2,07%) | 11 |
| Ménages unifamiliaux | 60 762 | * | (+0,87%) | 92 |
| Soutiens du ménage < 25 ans | -9 011 | * | (-1,98%) | 36 |
| Soutiens du ménage 25-34 ans | -2 409 | | (-0,11%) | 52 |
| Soutiens du ménage 35-44 ans | 7 652 | * | (+0,33%) | 62 |
| Soutiens du ménage 45-54 ans | 7 495 | * | (+0,46%) | 58 |
| Soutiens du ménage 55-64 ans | 685 | | (+0,05%) | 49 |
| Soutiens du ménage 65-74 ans | 1 265 | | (+0,11%) | 46 |
| Soutiens du ménage > 74 ans | -5 676 | * | (-0,76%) | 36 |
| Soutiens du ménage sexe masculin | -28 260 | * | (-0,41%) | 38 |
| Soutiens du ménage sexe féminin | 28 260 | * | (+0,95%) | 62 |

* Ces différences se sont révélées statistiquement significatives au niveau de 5 %.

Les groupes suivants avaient clairement tendance à être surreprésentés dans l'échantillon : les femmes, les groupes d'âges 0-4 ans, 5-9 ans, 10-14 ans et 45-49 ans, ainsi que les personnes dans les familles de recensement, particulièrement les personnes mariées et les enfants. Les groupes de personnes suivants étaient sous-représentés dans l'échantillon : les groupes d'âges 20-24 ans, 25-29 ans et, plus de 74 ans, les personnes divorcées et les personnes séparées, ainsi que les personnes hors famille de recensement.

En ce qui a trait aux caractéristiques des ménages, les logements possédés et les soutiens du ménage de sexe féminin avaient tendance à être surreprésentés dans l'échantillon tandis que les logements loués et les logements dont le type (p. ex. maison individuelle) était classé parmi les «autres logements» avaient tendance à être sous-représentés. Les ménages unifamiliaux et les familles époux-épouse tendaient à être surreprésentés tandis que les ménages non familiaux étaient sous-représentés. En outre, on a constaté une surreprésentation des ménages de trois, de quatre et de cinq personnes et une sous-représentation de ceux d'une personne. Les soutiens du ménage de 45-54 ans étaient surreprésentés, tandis que ceux de moins de 25 ans et de plus de 74 ans étaient sous-représentés.

Le tableau 2 montre que les moyennes des valeurs $Z^{(0)}$ (dans la colonne «Tous les enregistrements») s'écartaient de 0 d'une valeur supérieure à celle pouvant être attribuée à la variabilité d'échantillonnage, pour de nombreuses caractéristiques. Dans le cas des 43 valeurs moyennes marquées du symbole (*), l'hypothèse selon laquelle la moyenne des valeurs $Z^{(0)}$ est égale à zéro a été rejetée au niveau de 5 %. La colonne voisine de celle des moyennes, T: moy=0, donne la valeur de la statistique t pour le test de l'hypothèse selon laquelle la moyenne était égale à zéro[9]. Les autres colonnes du tableau 2 sont décrites dans les paragraphes qui suivent. Des graphiques montrant les histogrammes des valeurs $Z^{(0)}$ superposés à la distribution normale ont été produits pour deux caractéristiques, afin de donner une représentation visuelle des résultats du tableau 2. Ces graphiques, présentés à l'annexe D, visent la population totale et les soutiens du ménage de sexe masculin. Le graphique relatif à la population totale montre que l'histogramme est décalé vers la droite (moyenne=0,71) par rapport à la distribution normale. Le graphique relatif aux soutiens du ménage de sexe masculin montre que l'histogramme est décalé vers la gauche (moyenne=-0,32).

9 Ce test devrait être valide compte tenu du grand nombre d'observations (284 DR) et du degré élevé de normalité des valeurs Z(0) pour la plupart des caractéristiques.

**Tableau 2. Statistiques sommaires de 1991 pour les moyennes des valeurs $Z^{(0)}$ au niveau des DR (SD échantillonnés)**

| Caractéristiques étudiées | Tous les enregistrements | | Manquants/refus exclus (A) | | Conversions inversées (B) | | A et B | |
|---|---|---|---|---|---|---|---|---|
| | Moyenne | T:Moy=0 | Moyenne | T:Moy=0 | Moyenne | T:Moy=0 | Moyenne | T:Moy=0 |
| **Personnes** | | | | | | | | |
| Hommes | 0,32* | 5,08 | 0,19* | 3,04 | 0,14* | 2,14 | 0,16* | 2,41 |
| Femmes | 0,76* | 11,97 | 0,65* | 10,27 | 0,53* | 8,51 | 0,60* | 9,55 |
| Population totale | 0,71* | 11,20 | 0,55* | 8,55 | 0,44* | 6,66 | 0,49* | 7,60 |
| Âge, 0-4 | 0,32* | 5,21 | 0,27* | 4,43 | 0,19* | 3,14 | 0,25* | 4,01 |
| Âge, 5-9 | 0,50* | 7,61 | 0,44* | 6,80 | 0,39* | 6,04 | 0,42* | 6,43 |
| Âge, 10-14 | 0,50* | 8,24 | 0,44* | 7,24 | 0,40* | 6,58 | 0,42* | 6,87 |
| Âge, 15-19 | 0,21* | 3,41 | 0,14* | 2,35 | 0,13* | 2,12 | 0,13* | 2,13 |
| Âge, 20-24 | -0,23* | -3,60 | -0,29* | -4,53 | -0,29* | -4,45 | -0,28* | -4,46 |
| Âge, 25-29 | -0,15* | -2,52 | -0,21* | -3,41 | -0,23* | -3,70 | -0,22* | -3,57 |
| Âge, 30-34 | 0,21* | 3,18 | 0,16* | 2,45 | 0,14 | 2,08 | 0,15* | 2,24 |
| Âge, 35-39 | 0,19* | 3,00 | 0,14* | 2,19 | 0,11* | 1,77 | 0,12 | 1,90 |
| Âge, 40-44 | 0,25* | 4,30 | 0,20* | 3,52 | 0,19* | 3,32 | 0,19* | 3,30 |
| Âge, 45-49 | 0,26* | 4,31 | 0,22* | 3,72 | 0,21* | 3,60 | 0,21* | 3,48 |
| Âge, 50-54 | 0,14* | 2,27 | 0,12* | 2,06 | 0,12* | 2,03 | 0,12* | 2,03 |
| Âge, 55-59 | 0,00 | -0,06 | 0,01 | 0,19 | 0,02 | 0,28 | 0,02 | 0,28 |
| Âge, 60-64 | 0,05 | 0,99 | 0,08 | 1,46 | 0,07 | 1,25 | 0,08 | 1,47 |
| Âge, 65-74 | -0,01 | -0,18 | 0,04 | 0,71 | 0,04 | 0,63 | 0,04 | 0,60 |
| Âge, > 74 | -0,38* | 6,55 | -0,34* | -5,73 | -0,33* | -5,51 | -0,33* | -5,61 |
| Célibataires | 0,20* | 3,21 | 0,04 | 0,62 | -0,01 | -0,18 | 0,01 | 0,18 |
| Personnes mariées | 1,08* | 17,42 | 1,05* | 16,92 | 0,93* | 14,80 | 0,99* | 16,07 |
| Veufs et veuves | -0,36* | -6,11 | -0,30* | -5,04 | -0,24* | -4,02 | -0,27* | -4,48 |
| Personnes divorcées | -0,31* | -5,19 | -0,36* | -5,98 | -0,34* | -5,57 | -0,35* | -5,86 |
| Personnes séparées | -0,32* | -5,17 | -0,35* | -5,70 | -0,30* | -4,99 | -0,34* | -5,56 |
| **Familles de recensement** | | | | | | | | |
| Nombre total de familles | 1,20* | 18,90 | 1,01* | 16,48 | 0,76* | 12,27 | 0,91* | 14,99 |
| Familles époux-épouse | 1,18* | 18,10 | 1,02* | 16,14 | 0,77* | 12,24 | 0,93* | 14,92 |
| Familles monoparentales | -0,12* | -2,05 | -0,14* | -2,35 | -0,10 | -1,70 | -0,14* | -2,41 |
| Enfants dans les familles | 0,72* | 11,69 | 0,61* | 10,07 | 0,52* | 8,48 | 0,57* | 9,35 |
| Personnes dans les familles | 1,15* | 17,91 | 0,98* | 15,80 | 0,77* | 12,43 | 0,89* | 14,56 |
| Personnes hors famille | -1,47* | -18,79 | -1,37* | -16,85 | -1,07* | -13,82 | -1,28* | -16,15 |
| **Ménages et logements** | | | | | | | | |
| Logements possédés | 0,87* | 12,85 | 0,69* | 11,36 | 0,56* | 9,89 | 0,65* | 10,81 |
| Logements loués | -0,87* | -12,85 | -0,69* | -11,36 | -0,56* | -9,89 | -0,65* | -10,81 |
| Maisons individuelles | 0,57* | 10,63 | 0,36* | 7,55 | 0,23* | 5,07 | 0,33* | 6,95 |
| App. dans un immeuble ≥ 5 étages | 0,00 | 0,12 | 0,01 | 0,32 | 0,04 | 0,96 | 0,03 | 0,78 |
| Logements mobiles | -0,04 | -0,90 | -0,03 | -0,53 | -0,02 | -0,37 | -0,02 | -0,35 |
| Autres logements | -0,56* | -10,50 | -0,36* | -7,32 | -0,23* | -4,92 | -0,33* | -6,71 |
| Nombre total de ménages | 0,00 | - | 0,00 | - | 0,00 | - | 0,00 | - |
| Ménages d'une personne | -1,04* | -17,07 | -0,84* | -14,26 | -0,65* | -10,69 | -0,75* | -12,67 |
| Ménages de 2 personnes | 0,18* | 3,07 | 0,15* | 2,54 | 0,08 | 1,34 | 0,12* | 2,08 |
| Ménages de 3 personnes | 0,20* | 3,56 | 0,15* | 2,69 | 0,12* | 2,15 | 0,13* | 2,41 |
| Ménages de 4 personnes | 0,48* | 7,62 | 0,41* | 6,59 | 0,35* | 5,58 | 0,38* | 6,15 |
| Ménages de 5 personnes | 0,27* | 4,55 | 0,23* | 3,83 | 0,21* | 3,53 | 0,21* | 3,61 |
| Ménages de 6 personnes et plus | -0,08 | -1,36 | -0,12 | -1,93 | -0,12 | -1,92 | -0,13* | -2,05 |
| Ménages non familiaux | -1,31* | -20,42 | -1,13* | -18,17 | -0,87* | -14,16 | -1,03* | -16,81 |
| Ménages unifamiliaux | 1,37* | 20,99 | 1,19* | 18,86 | 0,94* | 15,19 | 1,10* | 17,64 |
| Soutiens du ménage < 25 ans | -0,32* | -5,69 | -0,35* | -5,98 | -0,32* | -5,44 | -0,33* | -5,71 |
| Soutiens du ménage 25-34 ans | 0,08 | 1,33 | 0,04 | 0,72 | 0,03 | 0,43 | 0,04 | 0,58 |
| Soutiens du ménage 35-44 ans | 0,27* | 4,40 | 0,23* | 3,69 | 0,22* | 3,53 | 0,22* | 3,50 |
| Soutiens du ménage 45-54 ans | 0,21* | 3,82 | 0,19* | 3,39 | 0,20* | 3,56 | 0,19* | 3,32 |
| Soutiens du ménage 55-64 ans | -0,04 | -0,80 | -0,02 | -0,28 | -0,01 | -0,25 | -0,01 | -0,13 |
| Soutiens du ménage 65-74 ans | -0,10 | -1,76 | -0,03 | -0,55 | -0,02 | -0,38 | -0,03 | -0,51 |
| Soutiens du ménage > 74 ans | -0,39* | -6,52 | -0,33* | -5,41 | -0,32* | -5,23 | -0,32* | -5,24 |
| Soutiens du ménage sexe masculin | -0,32* | -4,78 | -0,39* | -5,80 | -0,49* | -7,05 | -0,43* | -6,28 |
| Soutiens du ménage sexe féminin | 0,32* | 4,78 | 0,39* | 5,80 | 0,49* | 7,05 | 0,43* | 6,28 |

* Ces différences se sont révélées statistiquement significatives au niveau de 5 %.

## C. Explication du biais

Comme il a été mentionné plus haut, les différences observées entre les estimations-échantillon (basées sur une pondération simple) et les chiffres de population peuvent être expliquées de différentes façons. Il est possible qu'elles soient attribuables au fait qu'il y a eu 253 156 ménages (2,6 % du total) manquants ou ayant refusé de répondre dans le recensement de 1991. Il s'agissait soit de ménages ayant complètement refusé de répondre aux questions, soit de ménages pour lesquels le recenseur avait été incapable d'obtenir des renseignements (habituellement parce que les membres du ménage étaient absents au moment où le recensement a été mené ou avaient déménagé, le jour même du recensement ou après cette date, sans répondre au questionnaire). Le recenseur pouvait parfois déterminer le nombre de personnes et le mode d'occupation du logement, et inscrivait presque toujours le type de logement, mais en général toutes les autres réponses devaient être imputées pour ces ménages. De l'ensemble des ménages manquants ou ayant refusé de répondre, 43 155 étaient des ménages de l'échantillon. En outre, 6 753 ménages de l'échantillon, même s'ils n'étaient pas manquants ou n'avaient pas refusé de répondre (c.-à-d. qu'ils avaient fourni certaines réponses aux questions de base), n'ont fourni aucune réponse aux questions-échantillon. Au cours du traitement des données, ces 43 155 + 6 753 = 49 908 ménages de l'échantillon ayant une non-réponse complète pour les questions-échantillon ont été retranchés de l'échantillon (c.-à-d. qu'ils ont été convertis de ménages 2B en ménages 2A) et seules les réponses aux questions de base ont été imputées. Ce processus de conversion de ménages inclus dans l'échantillon en ménages non inclus dans l'échantillon est appelé conversion de document 2A/2B. Il est possible que les ménages manquants ou refusant de répondre, ainsi que les ménages n'ayant pas répondu aux questions-échantillon, aient des caractéristiques différentes des autres ménages (ils pourraient être, par exemple, de plus petite taille). Ainsi, la conversion de ménages 2B en ménages 2A pourrait biaiser l'échantillon. Par ailleurs, si le système d'imputation avait tendance à imputer plus souvent certaines caractéristiques aux ménages manquants ou refusant de répondre qu'à d'autres types de ménages, des divergences entre les estimations-échantillon et les chiffres de population auraient été produites, car seuls des ménages non inclus dans l'échantillon seraient touchés.

Pour examiner l'effet des ménages manquants ou refusant de répondre et de la conversion de document 2A/2B sur le biais d'échantillonnage, on a étudié trois situations différentes. Premièrement, les ménages manquants ou refusant de répondre ont été exclus (253 156 ménages), les coefficients de pondération simples ont été ajustés en conséquence et les statistiques $Z^{(0)}$ ont été calculées de nouveau (voir la colonne «Manquants/refus exclus» du tableau 2). Deuxièmement, plutôt que d'exclure les ménages manquants ou refusant de répondre, les documents 2A ont été reconvertis à leur état initial de documents 2B, afin qu'ils soient inclus dans l'échantillon (voir la colonne «Conversions inversées» du tableau 2). Cela a entraîné la reconversion de 49 908 ménages à l'état 2B. Enfin, la troisième situation incluait à la fois les reconversions et l'exclusion des ménages manquants ou refusant de répondre (voir la colonne «A et B» du tableau 2). Dans ce cas, 6 753 ménages ont été reconvertis à l'état 2B et 253 156 ménages ont été exclus. Le biais est demeuré statistiquement significatif au niveau de 5 % pour 42 des 53 caractéristiques après exclusion des ménages manquants ou refusant de répondre, pour 39 des 53 caractéristiques après les reconversions de documents, et pour 42 des 53 caractéristiques après aussi bien les reconversions de documents que l'exclusion des ménages manquants ou refusant de répondre. Ainsi, bien que ces facteurs aient clairement contribué au biais, une grande partie de ce dernier demeure. L'essentiel du biais restant est probablement attribuable à un ou plusieurs facteurs comme le biais de non-réponse, le biais de réponse, la sélection d'un échantillon biaisé par les recenseurs.

# VI. Évaluation des méthodes de pondération

## A. Formation des régions de pondération (RP)

La première étape du processus de pondération a été la formation des RP. Une RP est la plus petite unité géographique pour laquelle une concordance des caractéristiques de la population entre certains résultats-échantillon et chiffres de population peut être assurée. Une RP satisfait les conditions suivantes :

a) une RP devrait contenir entre 2 000 et 7 000 personnes (chiffre de population);

b) les limites des RP doivent respecter les limites des divisions de recensement (DR) et, autant que possible, des subdivisions de recensement (SDR), des secteurs de recensement (SR) et des circonscriptions électorales fédérales (CÉF);

c) les RP devraient être formées de SD contigus (c.-à-d. être continues).

Les SD échantillonnés ont été agencés en 5 736 RP comptant en moyenne 4 583 personnes (les personnes vivant dans des logements collectifs étant exclues). Sur les 5 736 RP, 5 727 (99,8 %) avaient une population se situant entre 2 000 et 7 000. Les neuf RP qui restaient avaient toutes une population inférieure à 2 000, car chacune était formée seulement d'une DR entière ayant une population de moins de 2 000. L'application de la PEGMC donnait des résultats acceptables pour seulement deux de ces neuf RP. Les sept RP restantes n'ont pas été soumises à la PEGMC. L'une a fait l'objet d'une pondération spéciale, tandis que les six autres ne contenaient aucun SD échantillonné et n'exigeaient par conséquent aucune pondération. La RP qui a fait l'objet d'une pondération spéciale au moyen d'un prototype de la PEGMC était formée d'une DR complète dont les SD échantillonnés contenaient une population de seulement 38 personnes.

On a examiné dans quelle mesure les RP respectaient les limites de chacun des différents types d'unités géographiques, soit les SR, les SDR dans les régions divisées en secteurs de recensement, les SDR dans les régions non divisées en secteurs de recensement et les CÉF. Étant donné que les limites des DR étaient toujours respectées, il n'a pas été nécessaire de faire de vérification à ce niveau. Seule la partie échantillonnée des unités géographiques a fait l'objet d'un examen visant à déterminer si les limites avaient été respectées. Les unités géographiques qui ne contenaient pas de SD échantillonnés ont été exclues de l'étude.

Le tableau 3 montre dans quelle mesure les RP ont respecté les limites des SR, des SDR et des CÉF. La première colonne montre le pourcentage d'unités géographiques ne contenant que des RP entières. La deuxième colonne indique le pourcentage d'unités géographiques qui étaient trop petites pour former des RP entières, mais étaient complètement englobées par une RP. La troisième colonne présente le pourcentage d'unités qui comprenaient des parties de RP différentes.

**Tableau 3. Mesure dans laquelle les régions de pondération respectaient les diverses limites géographiques**

| Unités géographiques | Contenant seulement des RP entières | Englobées entièrement par une RP différentes | Contenant des parties de RP |
|---|---|---|---|
| Divisions de recensement | 100 % | 0 % | 0 % |
| Secteurs de recensement | 58 % | 31 % | 11 % |
| Subdivisions de recensement dans des régions divisées en secteurs de recensement | 59 % | 32 % | 9 % |
| Subdivisions de recensement dans des régions non divisées en secteurs de recensement | 8 % | 84 % | 8 % |
| Circonscriptions électorales fédérales | 17 % | 0 % | 83 % |

Comme la PEGMC est appliquée indépendamment à l'intérieur des RP, la concordance entre les estimations-échantillon et les chiffres de population est assurée seulement dans le cas des unités géographiques contenant uniquement des RP entières. La concordance n'est pas assurée pour les unités géographiques qui sont complètement englobées par une RP ou qui contiennent des parties de RP différentes.

Les RP respectaient les limites des SR et des SDR des régions divisées en secteurs de recensement presque aussi bien dans un cas que dans l'autre. En raison de la petite taille de la plupart des SDR dans les régions non divisées en secteurs de recensement, la majorité d'entre elles (85 %) étaient englobées par une RP.

Seulement 17 % des CÉF ne contenaient que des RP entières. En outre, comme les CÉF sont considérablement plus grandes que les RP, aucune d'elles ne se trouvait complètement comprise dans une partie d'une RP. La majorité des CÉF (83 %) contenaient entre 5 et 30 RP entières, et entre 1 et 10 RP partielles. Leurs limites n'ont donc pas été bien respectées au moment de la formation des RP. Étant donné que les limites des CÉF n'ont aucun rapport avec celles des DR, des SR et des SDR, on ne pourrait pas mieux les respecter sans courir le risque d'enfreindre davantage les limites de ces dernières.

Pour de plus amples renseignements sur cette étude, voir Majkowski (1992b).

## B. Évaluation de la procédure d'estimation généralisée des moindres carrés

Le tableau qui suit présente la distribution, à l'échelle du Canada, du nombre de ménages de l'échantillon de 1991 dont le coefficient de pondération final selon la PEGMC se situait dans un intervalle particulier. Le tableau indique aussi la distribution de 1986, à l'échelle du Canada, pour les mêmes intervalles de coefficients de pondération. En 1986, des coefficients de pondération étaient attribués à la fois aux ménages et aux personnes en vertu de la MIQ. Ainsi, la colonne du tableau intitulée «1986 – MÉNAGES» donne la distribution du nombre de ménages de l'échantillon dont le coefficient de pondération de ménage se situait dans l'intervalle visé, tandis que la colonne «1986 – PERSONNES» donne la distribution du nombre de personnes dans les ménages de l'échantillon dont le coefficient de pondération de personnes appartenait à l'intervalle visé. Les pourcentages (arrondis à une décimale) par rapport au total de ménages de l'échantillon ou au total de personnes dans les ménages de l'échantillon au Canada pour les colonnes respectives sont également présentés au tableau 4.

**Tableau 4. Distribution des coefficients de pondération**

| Coefficient de pondération | 1991 | | 1986 – Ménages | | 1986 – Personnes | |
|---|---|---|---|---|---|---|
| | Nombre | % | Nombre | % | Nombre | % |
| < 1,00 | 0 | - | 44 | 0,0 | 136 | 0,0 |
| 1,00-1,99 | 58 934 | 3,1 | 245 | 0,0 | 1 041 | 0,0 |
| 2,00-2,99 | 170 352 | 8,9 | 8 534 | 0,5 | 25 825 | 0,5 |
| 3,00-3,99 | 338 334 | 17,7 | 178 362 | 10,3 | 415 754 | 8,7 |
| 4,00-4,99 | 432 603 | 22,6 | 683 232 | 39,6 | 1 946 279 | 40,9 |
| 5,00-5,99 | 369 274 | 19,3 | 578 431 | 33,5 | 1 676 221 | 35,2 |
| 6,00-6,99 | 242 000 | 12,7 | 180 378 | 10,5 | 483 424 | 10,2 |
| 7,00-7,99 | 137 472 | 7,2 | 56 111 | 3,3 | 133 233 | 2,8 |
| 8,00-8,99 | 74 270 | 3,9 | 20 755 | 1,2 | 44 168 | 0,9 |
| 9,00-9,99 | 39 671 | 2,1 | 9 001 | 0,5 | 16 954 | 0,4 |
| 10,00-14,99 | 45 808 | 2,4 | 8 348 | 0,5 | 15 033 | 0,3 |
| 15,00-19,99 | 3 182 | 0,2 | 700 | 0,0 | 1 230 | 0,0 |
| 20,00-24,99 | 357 | 0,0 | 66 | 0,0 | 148 | 0,0 |
| ≥ 25,00 | 0 | - | 40 | 0,0 | 38 | 0,0 |
| **Total** | **1 912 257** | **100,0** | **1 724 247** | **100,0** | **4 759 484** | **100,0** |

Le tableau ci-dessus montre que les coefficients de pondération PEGMC se situaient tous dans l'intervalle requis de 1 à 25. Pour toutes les régions (Est, Québec, Ontario et Ouest), la distribution de 1991 est semblable à la distribution à l'échelle nationale indiquée ci-dessus. La majorité des coefficients de pondération de 1991 sont centrés autour de 5, ce qui est normal dans une situation où un ménage sur cinq est inclus dans l'échantillon. Toutefois, comme l'indique le tableau 4, le pourcentage de ménages de l'échantillon de 1991 ayant un coefficient de pondération entre 3 et 7 est de 72,3 % seulement, tandis qu'il était de 93,9 % en 1986. Le pourcentage des personnes dans les ménages de l'échantillon ayant un coefficient de pondération de personne dans cet intervalle était de 95 % en 1986. Par ailleurs, en 1986, quelques coefficients de pondération étaient inférieurs à 1 ou supérieurs à 25.

Pour chacune des 5 729 RP de 1991, on a calculé la moyenne du nombre de contraintes au niveau des RP qui étaient retranchées. Sur les 62 contraintes au niveau des RP qui ont été soumises à la PEGMC, 7,1 en moyenne ont été retranchées à titre de contraintes faibles, 6,3 à titre de contraintes LD, 8,0 à titre de contraintes QLD et 6,9 à titre de contraintes causant des coefficients de pondération aberrants. Le total moyen de contraintes au niveau des RP qui ont été retranchées était donc de 28,3 par RP. En conséquence, pour chaque RP, il devrait y avoir un écart entre l'estimation et le chiffre de population pour un total moyen de 28,3 contraintes sur les 62 établies au niveau des RP. Il restait donc 31,7 contraintes par RP, en moyenne, ayant un écart de zéro. Pour chacune des RP, il y avait aussi des contraintes représentant des totaux au niveau des SD, plus précisément le nombre de ménages au niveau des SD (HHEACT) et la population totale au niveau des SD (PPEACT). En moyenne, pour chaque RP, HHEACT a été retranchée 1,3 fois et PPEACT a été retranchée 2,1 fois. Les RP, en moyenne, sont formées de 6,4 SD.

Le tableau 5 ci-dessous indique combien de fois une contrainte a été retranchée au niveau des RP, et pour quelle raison. Les contraintes sont rangées en ordre à partir des plus fréquemment retranchées jusqu'au moins fréquemment retranchées, pour les 5 729 RP. Les contraintes sont désignées par des abréviations qui sont définies à l'annexe A.

**Tableau 5. Fréquence d'exclusion des contraintes au niveau des RP en 1991**

| Contrainte | Faible | LD | QLD | Valeur aberrante | Total | Contrainte | Faible | LD | QLD | Valeur aberrante | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| FAMCHGE4 | 5109 | 609 | 11 | 0 | 5729 | AGEHM74 | 495 | 18 | 720 | 1256 | 2489 |
| AGEC1718 | 1496 | 4181 | 22 | 12 | 5711 | FAMCHLD0 | 1 | 24 | 1946 | 363 | 2334 |
| HHSIZEG6 | 3791 | 1636 | 257 | 18 | 5702 | AGE9 | 68 | 814 | 16 | 1070 | 1968 |
| AGEC1517 | 5257 | 64 | 302 | 43 | 5666 | HHSIZE4 | 56 | 149 | 841 | 912 | 1958 |
| MOVABLE | 4838 | 432 | 114 | 138 | 5522 | AGE49 | 3 | 117 | 231 | 1589 | 1940 |
| SEP | 1307 | 3892 | 86 | 226 | 5511 | AGE24 | 1 | 208 | 261 | 1401 | 1871 |
| AGEHM24 | 2939 | 1571 | 896 | 56 | 5462 | CHILD | 0 | 1237 | 290 | 269 | 1796 |
| HHSIZE1 | 176 | 371 | 4875 | 13 | 5435 | AGEHM64 | 53 | 34 | 313 | 1360 | 1760 |
| APT5PL | 4046 | 796 | 186 | 49 | 5077 | AGE74 | 200 | 0 | 270 | 1156 | 1626 |
| LONEPARF | 629 | 49 | 3725 | 618 | 5021 | AGEHM54 | 3 | 35 | 269 | 1120 | 1427 |
| AGEHM75P | 1690 | 618 | 2435 | 195 | 4938 | AGE44 | 1 | 5 | 84 | 1316 | 1406 |
| AGEC014 | 865 | 1 | 3497 | 574 | 4937 | AGE29 | 2 | 44 | 244 | 1078 | 1368 |
| FAMCHLD3 | 588 | 2136 | 1626 | 601 | 4851 | SINGDET | 316 | 548 | 344 | 147 | 1355 |
| AGEC617 | 3315 | 12 | 625 | 655 | 4607 | AGEHM34 | 11 | 8 | 297 | 859 | 1175 |
| NONMEMB | 1 | 4463 | 38 | 14 | 4516 | AGE39 | 1 | 5 | 54 | 1036 | 1096 |
| AGE75PL | 1108 | 2369 | 626 | 323 | 4426 | AGECLE17 | 7 | 1 | 746 | 310 | 1064 |
| AGECGE18 | 41 | 41 | 3726 | 608 | 4416 | HHSIZE2 | 1 | 2 | 580 | 461 | 1044 |
| FAMCHLD1 | 2 | 20 | 3804 | 255 | 4081 | AGE34 | 1 | 25 | 30 | 896 | 952 |
| HHSIZE5 | 515 | 1013 | 845 | 1288 | 3661 | FAMCHLD2 | 50 | 113 | 120 | 546 | 829 |
| AGE64 | 242 | 446 | 1781 | 966 | 3435 | AGEHM44 | 1 | 0 | 4 | 620 | 625 |
| OTHDWLS | 577 | 1519 | 1012 | 261 | 3369 | MARRIED | 0 | 1 | 63 | 266 | 330 |
| AGE4 | 34 | 2546 | 6 | 687 | 3273 | CENFAM | 0 | 0 | 243 | 55 | 298 |
| AGE54 | 12 | 478 | 1578 | 1185 | 3253 | OWNED | 22 | 1 | 60 | 200 | 283 |
| AGE14 | 75 | 2286 | 4 | 739 | 3104 | SINGLE | 0 | 0 | 7 | 268 | 275 |
| HHSIZE3 | 3 | 66 | 2189 | 670 | 2928 | HUSBAND | 0 | 28 | 54 | 116 | 198 |
| AGE59 | 74 | 496 | 701 | 1526 | 2797 | MALEGE15 | 0 | 0 | 104 | 51 | 155 |
| WIDOWED | 254 | 386 | 155 | 1857 | 2652 | MALEHM | 0 | 0 | 0 | 85 | 85 |
| AGECLE5 | 71 | 0 | 609 | 1935 | 2615 | MALE | 0 | 0 | 0 | 11 | 11 |
| DIVORCED | 185 | 42 | 244 | 2124 | 2595 | TOTPERS | 0 | 0 | 0 | 5 | 5 |
| AGE19 | 18 | 409 | 894 | 1267 | 2588 | TPERGE15 | 0 | 0 | 0 | 5 | 5 |
| AGEC614 | 96 | 0 | 607 | 1847 | 2550 | TOTHHLD | 0 | 0 | 0 | 0 | 0 |

Le tableau 5 montre que quatre des cinq contraintes les plus fréquemment retranchées dans l'ensemble (FAMCHGE4, HHSIZEG6, AGEC1517 ET MOVABLE) étaient quatre des cinq contraintes les plus fréquemment retranchées à titre de faibles contraintes. Ces quatre contraintes ont été retranchées parce qu'elles étaient faibles dans au moins 3 791 RP (ou 66 % des RP). Les contraintes s'appliquant au plus grand nombre de personnes, de ménages et de familles de recensement (p. ex. TOTHHLD, TOTPERS, MALE et CENFAM) faisaient partie des contraintes les moins fréquemment retranchées parce qu'elles étaient faibles ou pour les autres raisons. Les personnes hors famille de recensement (NONMEMB), les familles de recensement ayant certains enfants à la maison âgés de moins de 18 ans et dont les autres enfants ont 18 ans ou plus (AGEC1718), ainsi que les personnes séparées (SEP) sont les contraintes qui étaient les plus fréquemment retranchées à titre de contraintes LD. AGEC1718 a finalement été l'une des contraintes les plus fréquemment retranchées dans l'ensemble, car dans de nombreux RP, elle a aussi été retranchée parce qu'elle était faible. Les contraintes fréquemment retranchées à titre de contraintes QLD étaient les ménages d'une personne (HHSIZE1), les familles de recensement sans enfants à la maison (FAMCHLD0), les mères seules (LONEPARF) et les familles de recensement ayant à la maison tous leurs enfants, âgés de 18 ans ou plus (AGECGE18). Par ailleurs, les personnes divorcées (DIVORCED), les familles de recensement ayant à la maison tous leurs enfants, âgés de 5 ans ou moins (AGECLE5) et les veufs et veuves (WIDOWED) ont été les contraintes les plus fréquemment retranchées pour avoir donné des coefficients de pondération aberrants. La contrainte de la population totale (TOTPERS), qui a été retranchée dans quelques RP, a causé au départ une certaine inquiétude. Dans les tests de la PEGMC, la contrainte TOTPERS n'avait jamais été

retranchée, de sorte qu'il y avait toujours une concordance parfaite entre l'estimation de la population totale et le chiffre de population. Toutefois, au moment du traitement de production, TOTPERS a été retranchée dans cinq RP pour avoir entraîné des coefficients aberrants, et il en est résulté une légère sous-estimation de 150 personnes. Des modifications de la PEGMC permettant de conserver la contrainte TOTPERS dans ces RP ont été testées. Toutefois, puisque plusieurs autres contraintes importantes étaient retranchées au cours de ces tests et que la sous-estimation était relativement faible pour la population totale, il a été décidé de laisser TOTPERS exclu pour ces cinq RP.

Un des buts de la méthode de pondération consiste à réduire les écarts entre les chiffres de population et les estimations-échantillon correspondantes pour les contraintes. Ces écarts sont le résultat de la variabilité et du biais d'échantillonnage (voir le chapitre V). Certains écarts peuvent toutefois subsister même après l'application de la méthode de pondération. Les écarts correspondent à la différence entre les estimations-échantillon et les chiffres de population, laquelle est exprimée en tant que pourcentage des chiffres de population, c'est-à-dire :

$$\text{écart} = \frac{\text{estimation-échantillon} - \text{chiffre de population}}{\text{chiffre de population}} \times 100 \qquad (2)$$

Le numérateur de cette expression (estimation-échantillon – chiffre de population) est souvent appelé la «différence».

Le tableau 6 montre les différences (DIFF) et les écarts (ÉCART) à l'échelle du Canada, en 1991, pour les 62 contraintes. À noter que les valeurs ÉCART ont été arrondies à deux décimales. Toutes ces caractéristiques sont des contraintes au niveau des RP qui ont servi à déterminer les coefficients PEGMC en 1991. Les estimations-échantillon et les chiffres de population se fondent sur les logements privés occupés dans les SD échantillonnés. Dans ce tableau, les contraintes ont les mêmes abréviations que celles qui ont été utilisées au tableau 5, et qui sont définies à l'annexe A.

On peut voir au tableau 6 que quatre des 62 contraintes affichent un ÉCART de 0,00. Ces quatre contraintes correspondent aux quatre les moins fréquemment retranchées selon le tableau 5. Il ressort du tableau 6 que 24 contraintes ont entraîné une sous-estimation du chiffre de population, tandis que 34 ont entraîné une surestimation du chiffre de population pour la contrainte. Les contraintes présentant les sous-estimations les plus fortes en pourcentage étaient HHSIZEG6 (-3,21), FAMCHGE4 (-2,71) et MOVABLE (-0,78), tandis que les contraintes présentant les surestimations les plus fortes étaient AGEC617 (1,69), FAMCHLD3 (1,52) et HHSIZE5 (1,26). Toutes ces contraintes associées à une sous-estimation ou une surestimation élevée ne visent pas une grande proportion de la population, et elles sont parmi les contraintes les plus fréquemment retranchées selon le tableau 5.

## Tableau 6. Écarts estimation/population à l'échelle du Canada en 1991

| Contrainte | DIFF | Écart (%) | Contrainte | DIFF | Écart (%) | Contrainte | DIFF | Écart (%) |
|---|---|---|---|---|---|---|---|---|
| TOTPERS | -150 | 0,00 | DIVORCED | -4131 | -0,33 | AGEHM24 | 581 | 0,13 |
| TPERGE15 | -135 | 0,00 | WIDOWED | -6695 | -0,55 | AGEHM34 | 111 | 0,01 |
| MALE | -396 | 0,00 | SEP | 3708 | 0,64 | AGEHM44 | 3430 | 0,15 |
| MALEGE15 | -1022 | -0,01 | CENFAM | -438 | -0,01 | AGEHM54 | 5857 | 0,36 |
| AGE4 | -2151 | -0,12 | NONMEMB | -9916 | -0,24 | AGEHM64 | 1582 | 0,12 |
| AGE9 | -1789 | -0,10 | HUSBAND | 630 | 0,01 | AGEHM74 | -6122 | -0,53 |
| AGE14 | 3925 | 0,21 | CHILD | 9574 | 0,11 | AGEHM75P | -5439 | -0,73 |
| AGE19 | 8705 | 0,48 | LONEPARF | 1927 | 0,25 | FAMCHLD0 | -8031 | -0,31 |
| AGE24 | 4890 | 0,26 | TOTHHLD | 0 | 0,00 | FAMCHLD1 | -1874 | -0,10 |
| AGE29 | -8762 | -0,38 | OWNED | 1039 | 0,02 | FAMCHLD2 | 4637 | 0,24 |
| AGE34 | 580 | 0,02 | MALEHM | -1616 | -0,02 | FAMCHLD3 | 10277 | 1,52 |
| AGE39 | -3777 | -0,17 | SINGDET | 316 | 0,01 | FAMCHGE4 | -5435 | -2,71 |
| AGE44 | 1278 | 0,06 | MOVABLE | -1358 | -0,78 | AGECHLE5 | -1163 | -0,12 |
| AGE49 | 2665 | 0,17 | APT5PL | -313 | -0,03 | AGEC614 | -4299 | -0,45 |
| AGE54 | 3122 | 0,24 | OTHDWLS | 1354 | 0,04 | AGEC1517 | -1418 | -0,65 |
| AGE59 | 1639 | 0,14 | HHSIZE1 | -14571 | -0,65 | ACEC014 | 1524 | 0,26 |
| AGE64 | 1005 | 0,09 | HHSIZE2 | 3250 | 0,10 | AGEC617 | 5497 | 1,69 |
| AGE74 | -4312 | -0,24 | HHSIZE3 | 6227 | 0,36 | AGECLE17 | 892 | 0,03 |
| AGE75P | -7169 | -0,68 | HHSIZE4 | 6158 | 0,35 | AGECGE18 | 3976 | 0,35 |
| MARRIED | 4927 | 0,04 | HHSIZE5 | 9029 | 1,26 | AGEC1718 | 2739 | 0,56 |
| SINGLE | 2041 | 0,02 | HHSIZEG6 | -10092 | -3,21 | | | |

Une étude a été faite pour comparer les différences absolues entre les estimations-échantillon et les chiffres de population pour 62 caractéristiques observées en 1991 et en 1986, à différents niveaux géographiques. Les 62 caractéristiques incluses dans cette étude des différences absolues sont présentées à l'annexe B. Les résultats de l'étude sont résumés au tableau 7 ci-dessous. Le tableau indique le pourcentage de caractéristiques ayant une «valeur R» se situant à l'intérieur d'un certain intervalle pour les six niveaux géographiques visés. Une valeur R est un quotient entre les différences de 1991 et celles de 1986 calculé selon l'expression suivante :

$$R = 100 * \frac{\sum_{1}^{N_{91}} |\hat{X}^{91} - X^{91}| \; / \; \sum_{1}^{N_{91}} X^{91}}{\sum_{1}^{N_{86}} |\hat{X}^{86} - X^{86}| \; / \; \sum_{1}^{N_{86}} X^{86}} \qquad (3)$$

où $X^{91}$ and $X^{86}$ sont respectivement les chiffres de population de 1991 et de 1986 pour une caractéristique. L'estimation-échantillon de 1991 d'après les coefficients de pondération PEGMC est $\hat{X}^{91}$, tandis que l'estimation-échantillon de 1986 fondée sur les coefficients de pondération MIQ est $\hat{X}^{86}$. Des valeurs R ont été calculées pour chacun des six niveaux géographiques (SD, RP, SDR, DR, Prov. et Canada). La somme des valeurs absolues des différences population/estimation a été calculée, $N_{91}$ étant égal au nombre d'unités du niveau géographique visé en 1991 et $N_{86}$ étant égal au nombre d'unités du niveau géographique visé en 1986. Une valeur R se situant entre 95 et 105 indique que les systèmes d'estimation de 1991 et de 1986 ont donné à peu près le même résultat. Une valeur R inférieure à 95 indique que le système de 1991 a été plus efficace que celui de 1986 pour la caractéristique visée au niveau géographique en question, tandis qu'une valeur R supérieure à 105 indique qu'il a été moins efficace. Le tableau 7 donne aussi, pour les 37 caractéristiques des personnes et les 25 caractéristiques des ménages, le pourcentage se situant dans chacun des trois intervalles de valeurs R. Pour de plus amples renseignements sur cette étude, voir Majkowski (1994).

**Tableau 7. Pourcentage des caractéristiques ayant des valeurs R dans certains intervalles**

| Caractéristiques | Valeur R | SD | RP | SDR | DR | Prov. | Canada |
|---|---|---|---|---|---|---|---|
| Personnes (37) | < 95 | 84 | 51 | 76 | 41 | 22 | 22 |
| | 95-105 | 16 | 14 | 10 | 16 | 19 | 14 |
| | > 105 | 0 | 35 | 14 | 43 | 59 | 65 |
| Ménages (25) | < 95 | 92 | 68 | 88 | 56 | 44 | 40 |
| | 95-105 | 4 | 8 | 4 | 20 | 8 | 4 |
| | > 105 | 4 | 24 | 8 | 24 | 48 | 56 |
| Toutes (62) | < 95 | 87 | 58 | 81 | 47 | 31 | 29 |
| | 95-105 | 11 | 11 | 8 | 18 | 14 | 10 |
| | > 105 | 2 | 31 | 11 | 35 | 55 | 61 |

Le tableau 7 montre que sur l'ensemble des 62 caractéristiques, 87 % avaient une valeur R inférieure à 95 % au niveau des SD, tandis que seulement 2 % (c'est-à-dire une seule caractéristique) avaient une valeur R supérieure à 105 % au niveau des SD. Il s'ensuit que le système d'estimation de 1991 a réussi à réduire les différences population/estimation au niveau des SD, par rapport au système d'estimation de 1986. Toutefois, comme le montre le tableau, cette efficacité du système d'estimation de 1991 s'atténue constamment à mesure que s'élargit le niveau géographique. Aux niveaux des provinces et du Canada, le pourcentage de caractéristiques ayant une valeur R supérieure à 105 est de plus de 50 %. On peut observer une tendance semblable si l'on examine séparément les caractéristiques des personnes et les caractéristiques des ménages. La comparaison des deux ensembles de caractéristiques dans le tableau révèle que le pourcentage des caractéristiques qui ont une valeur R supérieure à 105 est plus faible pour les caractéristiques des ménages que pour les caractéristiques des personnes. Il en est ainsi à tous les niveaux géographiques, sauf au niveau des SD, mais la différence, dans ce cas, n'est que de 4 % (c'est-à-dire une seule caractéristique des ménages).

Les résultats du tableau 7 indiquent que les différences positives et négatives obtenues au niveau des RP ne s'annulent pas aussi bien en 1991 qu'en 1986 quand on fait la somme de ces différences à des niveaux géographiques supérieurs (DR, Prov. et Canada). Les différences de 1991, en pourcentage, à ces niveaux géographiques supérieurs, demeurent toutefois très faibles.

# VII. Examen de la cohérence entre les estimations-échantillonet les chiffres de population

Pour que la PEGMC soit efficace, il a fallu retrancher certaines contraintes dans chaque RP avant de calculer les coefficients de pondération. Plusieurs caractéristiques importantes ont donc été exclues dans un certain nombre de RP. Cela a eu pour conséquence de réduire le niveau de concordance (cohérence) entre les estimations-échantillon et les chiffres de population pour ces caractéristiques. De plus, de nombreuses unités géographiques d'intérêt ne se composent pas toujours de RP complètes (voir le chapitre VI, section A). Dans ces régions, la cohérence pour l'ensemble des caractéristiques dépend donc de la mesure dans laquelle les régions se composent de RP complètes.

L'examen de la cohérence a porté sur les écarts entre les estimations-échantillon et les chiffres de population (exprimés en pourcentage des chiffres de population) pour le même ensemble de base de 53 caractéristiques que celles de l'examen du biais d'échantillonnage (voir le chapitre V) et pour les unités géographiques suivantes :

a) divisions de recensement;

b) subdivisions de recensement;

c) secteurs de recensement et secteurs de recensement provinciaux;

d) secteurs de dénombrement.

L'annexe B présente la liste des caractéristiques dont les écarts sont étudiés dans ce chapitre. Comme il est indiqué au chapitre VI, section B, les écarts entre les estimations-échantillon et les chiffres de population ont été calculés de la façon suivante :

$$\text{écart} = \frac{\text{estimation-échantillon} - \text{chiffre de population}}{\text{chiffre de population}} \times 100$$

## A. Divisions de recensement (DR)

Les centiles du tableau 8 résument les résultats relatifs au niveau de cohérence, pour l'ensemble des 284 DR échantillonnées au Canada, d'une grande variété de caractéristiques de base dont le chiffre de population[10] est supérieur à 50. En général, les écarts obtenus pour des caractéristiques dont le chiffre de population était de 50 ou moins dans la plupart des unités géographiques se sont révélés relativement élevés (soit positifs soit négatifs). Il a donc été décidé de ne pas inclure les unités géographiques où le chiffre de population pour la caractéristique était inférieur ou égal à 50, car il suffisait de quelques-unes de ces unités pour perturber sensiblement les centiles des écarts présentés aux tableaux 8 à 12. Cela se produirait si une grande partie de ces unités présentaient soit des écarts positifs relativement élevés, soit des écarts négatifs relativement élevés. Au tableau 8, pour chaque caractéristique, N % des DR présentaient des écarts inférieurs au N$^{e}$ centile tandis que 100 - N % des DR avaient des écarts supérieurs au N$^{e}$ centile. L'écart se situait donc entre le 10$^{e}$ centile et le 90$^{e}$ centile pour 80 % des DR, entre le 25$^{e}$ et le 75$^{e}$ centile pour 50 % des DR, etc. Par exemple, l'écart pour le groupe d'âge 0-4 variait entre -2,86 % et 2,38 % pour 80 % des DR.

[10] Il s'agit du chiffre de population pour la caractéristique. Par exemple, le niveau de cohérence pour le groupe d'âge 0-4 est résumé pour toutes les DR dans lesquelles plus de 50 personnes appartenaient au groupe d'âge 0-4. La même définition vaut pour les tableaux 9, 10, 11 et 12.

**Tableau 8. Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les DR – recensements de 1991 et de 1986**

| Caractéristiques étudiées | Centiles des écarts - 1991 | | | | | 1986 | |
|---|---|---|---|---|---|---|---|
| | 10e | 25e | 50e | 75e | 90e | 10e | 90e |
| **Personnes** | | | | | | | |
| Hommes | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Femmes | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Population totale | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Âge, 0-4 | -2,86 | -1,07 | 0,00 | 1,06 | 2,38 | 0,00 | 0,00 |
| Âge, 5-9 | -2,14 | -1,05 | 0,00 | 0,63 | 2,06 | 0,00 | 0,00 |
| Âge, 10-14 | -1,80 | -0,56 | 0,00 | 0,99 | 2,36 | 0,00 | 0,00 |
| Âge, 15-19 | -1,87 | -0,51 | 0,55 | 1,76 | 3,15 | -0,88 | 1,02 |
| Âge, 20-24 | -3,71 | -0,95 | 0,32 | 2,35 | 4,14 | -1,35 | 1,15 |
| Âge, 25-29 | -3,07 | -1,40 | -0,20 | 0,20 | 1,78 | -4,12 | 3,52 |
| Âge, 30-34 | -1,67 | -0,39 | 0,00 | 0,62 | 2,11 | -3,58 | 3,73 |
| Âge, 35-39 | -3,14 | -0,84 | 0,00 | 0,59 | 2,26 | -4,03 | 3,46 |
| Âge, 40-44 | -2,33 | -0,69 | 0,00 | 1,02 | 3,04 | -5,07 | 5,89 |
| Âge, 45-49 | -2,95 | -1,14 | 0,00 | 1,80 | 4,26 | -4,82 | 4,98 |
| Âge, 50-54 | -4,96 | -2,14 | 0,13 | 2,02 | 5,03 | -5,43 | 5,17 |
| Âge, 55-59 | -6,13 | -2,33 | 0,00 | 1,60 | 4,21 | -4,47 | 5,69 |
| Âge, 60-64 | -3,69 | -1,75 | 0,07 | 1,93 | 4,88 | -5,28 | 4,91 |
| Âge, 65-74 | -2,28 | -1,08 | 0,00 | 0,59 | 2,03 | -2,55 | 4,06 |
| Âge, > 74 | -7,87 | -3,66 | -1,07 | 0,67 | 4,65 | -7,49 | 5,19 |
| Célibataires | -0,08 | 0,00 | 0,00 | 0,00 | 0,12 | -0,28 | 0,29 |
| Personnes mariées | 0,00 | 0,00 | 0,00 | 0,08 | 0,33 | -0,33 | 0,20 |
| Veufs et veuves | -4,22 | -2,26 | -0,55 | 0,57 | 2,27 | -4,49 | 5,55 |
| Personnes divorcées | -4,47 | -1,82 | -0,14 | 1,88 | 4,73 | -8,43 | 9,46 |
| Personnes séparées | -9,33 | -3,96 | 0,44 | 4,90 | 11,12 | -10,60 | 10,61 |
| **Familles de recensement** | | | | | | | |
| Nombre total de familles | -0,01 | 0,00 | 0,00 | 0,00 | 0,00 | -0,13 | 0,08 |
| Familles époux-épouse | -0,01 | 0,00 | 0,00 | 0,00 | 0,07 | -0,13 | 0,10 |
| Familles monoparentales | -0,80 | 0,00 | 0,00 | 0,00 | 0,16 | -0,14 | 0,04 |
| Enfants dans les familles | -0,08 | 0,00 | 0,03 | 0,21 | 0,42 | -0,07 | 0,21 |
| Personnes dans les familles | -0,04 | 0,00 | 0,02 | 0,09 | 0,18 | 0,00 | 0,02 |
| Personnes hors famille | -1,42 | -0,63 | -0,14 | 0,00 | 0,32 | -0,11 | 0,03 |
| **Ménages et logements** | | | | | | | |
| Logements possédés | -0,05 | 0,00 | 0,00 | 0,00 | 0,09 | 0,00 | 0,00 |
| Logements loués | -0,20 | 0,00 | 0,00 | 0,00 | 0,08 | 0,00 | 0,00 |
| Maisons individuelles | -0,07 | 0,00 | 0,00 | 0,00 | 0,07 | -0,53 | 0,50 |
| App. dans un immeuble ≥ 5 étages | -6,68 | -1,24 | 0,00 | 1,01 | 4,29 | -5,36 | 8,49 |
| Logements mobiles | -10,64 | -3,85 | -0,95 | 1,22 | 6,02 | -11,67 | 13,54 |
| Autres logements | -0,78 | -0,18 | 0,11 | 0,67 | 1,84 | -3,48 | 2,39 |
| Nombre total de ménages | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Ménages d'une personne | -2,54 | -1,44 | -0,60 | 0,03 | 0,73 | 0,00 | 0,00 |
| Ménages de 2 personnes | -0,51 | -0,04 | 0,00 | 0,22 | 0,87 | -1,64 | 2,17 |
| Ménages de 3 personnes | -2,43 | -0,92 | 0,21 | 1,23 | 3,17 | -4,36 | 3,95 |
| Ménages de 4 personnes | -1,66 | -0,57 | 0,04 | 0,92 | 2,15 | -3,33 | 4,17 |
| Ménages de 5 personnes | -3,83 | -0,59 | 1,56 | 4,19 | 6,79 | -7,78 | 7,22 |
| Ménages de 6 personnes et plus | -14,49 | -8,69 | -4,20 | 0,18 | 5,21 | -11,41 | 7,37 |
| Ménages non familiaux | -1,18 | -0,67 | -0,30 | 0,03 | 0,45 | 0,00 | 0,00 |
| Ménages unifamiliaux | -0,28 | -0,03 | 0,19 | 0,39 | 0,69 | -0,23 | 0,31 |
| Soutiens du ménage < 25 ans | -9,54 | -3,32 | -0,09 | 5,03 | 12,64 | -7,78 | 6,76 |
| Soutiens du ménage 25-34 ans | -1,40 | -0,35 | 0,00 | 0,47 | 1,48 | -1,75 | 1,66 |
| Soutiens du ménage 35-44 ans | -0,48 | 0,00 | 0,00 | 0,38 | 1,43 | -1,98 | 1,80 |
| Soutiens du ménage 45-54 ans | -1,81 | -0,28 | 0,14 | 1,11 | 3,41 | -2,53 | 3,16 |
| Soutiens du ménage 55-64 ans | -2,99 | -1,11 | 0,00 | 1,05 | 2,76 | -2,63 | 2,99 |
| Soutiens du ménage 65-74 ans | -2,91 | -1,47 | -0,34 | 0,51 | 2,64 | -3,75 | 5,03 |
| Soutiens du ménage > 74 ans | -8,09 | -4,09 | -1,00 | 1,30 | 4,77 | -8,94 | 5,78 |
| Soutiens du ménage sexe masculin | -0,04 | 0,00 | 0,00 | 0,00 | 0,00 | -0,74 | 0,42 |
| Soutiens du ménage sexe féminin | 0,00 | 0,00 | 0,00 | 0,00 | 0,12 | -1,56 | 2,73 |

Toutes les DR sont formées de RP complètes. Ainsi, parmi les caractéristiques qui se sont traduites en contraintes en 1991, celles qui étaient rarement ou jamais retranchées dans une RP affichaient une cohérence quasi parfaite au niveau des DR[11]. Ces caractéristiques étaient les suivantes : hommes, population totale, célibataires, personnes mariées, nombre total de familles de recensement, familles époux-épouse, logements possédés, total des ménages et soutiens du ménage de sexe masculin. Comme l'indique le tableau 5, ce groupe de caractéristiques n'a pas été retranché dans beaucoup de RP. Les personnes mariées, caractéristique la plus fréquemment retranchée dans ce groupe, ont été retranchées dans 330 RP (environ 5,8 % de l'ensemble de RP) par le système d'estimation. Le niveau de cohérence observé pour les autres caractéristiques n'était pas parfait, mais demeurait quand même très bon, sauf pour les caractéristiques représentant seulement un faible pourcentage de la population dans la plupart des DR, par exemple, les personnes séparées, les logements mobiles, les ménages de 6 personnes ou plus et les soutiens du ménage âgés de moins de 25 ans. Une relation générale existe entre les écarts et les chiffres de population pour toutes les caractéristiques, c'est-à-dire que la cohérence s'accroît à mesure qu'augmente le chiffre de population pour la DR.

Les deux dernières colonnes du tableau 8 indiquent le 10$^{e}$ et le 90$^{e}$ centile des écarts obtenus en 1986 pour les DR. Les écarts de 1986 se fondent évidemment sur les estimations-échantillon résultant de la MIQ, qui était la méthode appliquée au recensement de 1986. Les tableaux 9, 10, 11 et 12 qui suivent présentent aussi ces deux colonnes pour les autres niveaux géographiques. La comparaison avec les mêmes centiles en 1991 permet d'observer que les écarts de 1991 au niveau des DR sont semblables ou sensiblement inférieurs aux écarts de 1986 pour deux tiers des caractéristiques du tableau 8. La taille des écarts au niveau des DR est très faible en comparaison des écarts observés aux autres niveaux géographiques, qui sont traités dans les sections qui suivent.

## B. Subdivisions de recensement (SDR)

Le tableau 9 résume les résultats concernant la cohérence entre les estimations-échantillon et les chiffres de population pour toutes les SDR échantillonnées au Canada dont le chiffre de population pour la caractéristique était supérieur à 50. Il porte sur les mêmes caractéristiques que le tableau 8. Les SDR ne sont pas toujours uniquement formées de RP entières. En outre, elles sont beaucoup plus petites en moyenne que les DR. Par conséquent, le niveau de cohérence n'est pas aussi bon pour les SDR que pour les DR. De façon générale, comme dans le cas des DR, la cohérence devenait meilleure à mesure qu'augmentaient les chiffres de population dans les SDR, et ce, pour toutes les caractéristiques. Comparativement aux écarts de 1986 au 10$^{e}$ et au 90$^{e}$ centile, les écarts de 1991 sont beaucoup plus faibles pour un grand nombre de caractéristiques.

## C. Secteurs de recensement (SR) et secteurs de recensement provinciaux (SRP)

Le tableau 10 résume les résultats relatifs au niveau de cohérence pour tous les SR échantillonnés au Canada, et le tableau 11 fait de même pour tous les SRP échantillonnés au Canada. Les deux tableaux ne tiennent compte que des SR et des SRP ayant des chiffres de population supérieurs à 50 pour la caractéristique. Les SR et les SRP sont en moyenne plus populeux que les SDR. Les SRP sont légèrement plus populeux en moyenne que les SR, mais les limites des SR ont été mieux respectées que celles des SRP au moment de la formation des RP. Par conséquent, tant en 1991 qu'en 1986, la cohérence était meilleure au niveau des SR qu'au niveau des SRP pour la plupart des caractéristiques tandis qu'elle était meilleure au niveau des SRP qu'au niveau des SDR pour la plupart des caractéristiques. Les caractéristiques faisant exception à cette règle étaient généralement celles qui présentaient une cohérence médiocre à tous les niveaux géographiques. En comparaison des écarts de 1986 au 10$^{e}$ et au 90$^{e}$ centile, les écarts de 1991 sont ici encore beaucoup plus faibles pour un grand nombre de caractéristiques, tant au niveau des SR qu'à celui des SRP.

---

11 Même lorsque les caractéristiques présentent une cohérence parfaite, les totalisations publiées des caractéristiques de base établies à partir des données-échantillon ne concordent pas parfaitement avec les totalisations des mêmes caractéristiques basées sur les données intégrales. Cette différence est attribuable au fait que les résidents des logements collectifs qui n'avaient pas à répondre aux questions-échantillon (voir le chapitre II, section B) sont inclus dans les totalisations faites à partir des données intégrales, mais exclus des totalisations fondées sur les données-échantillon.

**Tableau 9. Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SDR – recensements de 1991 et de 1986**

| Caractéristiques étudiées | Centiles des écarts – 1991 | | | | | 1986 | |
|---|---|---|---|---|---|---|---|
| | 10e | 25e | 50e | 75e | 90e | 10e | 90e |
| **Personnes** | | | | | | | |
| Hommes | -5,47 | -1,69 | 0,00 | 1,62 | 5,05 | -9,15 | 9,76 |
| Femmes | -5,49 | -1,83 | 0,00 | 1,83 | 5,56 | -9,38 | 8,97 |
| Population totale | -3,36 | -0,37 | 0,00 | 0,26 | 3,28 | -7,56 | 7,95 |
| Âge, 0-4 | -19,44 | -7,14 | 0,00 | 6,23 | 18,50 | -19,36 | 17,63 |
| Âge, 5-9 | -16,53 | -6,11 | 0,00 | 5,73 | 17,01 | -19,17 | 18,71 |
| Âge, 10-14 | -16,99 | -6,15 | 0,00 | 7,06 | 18,01 | -19,24 | 18,92 |
| Âge, 15-19 | -16,79 | -6,03 | 0,00 | 6,92 | 19,04 | -21,32 | 20,48 |
| Âge, 20-24 | -20,98 | -7,71 | 0,00 | 7,94 | 21,04 | -20,41 | 20,18 |
| Âge, 25-29 | -20,34 | -7,54 | 0,00 | 5,96 | 19,29 | -21,36 | 20,59 |
| Âge, 30-34 | -17,90 | -6,21 | 0,00 | 6,02 | 17,39 | -21,07 | 21,74 |
| Âge, 35-39 | -19,22 | -6,66 | 0,00 | 5,86 | 18,71 | -20,97 | 20,66 |
| Âge, 40-44 | -19,19 | -6,43 | 0,00 | 7,42 | 19,92 | -22,08 | 21,03 |
| Âge, 45-49 | -19,21 | -7,37 | 0,00 | 8,45 | 22,15 | -20,62 | 21,18 |
| Âge, 50-54 | -23,11 | -9,06 | 0,00 | 8,62 | 21,29 | -21,80 | 22,17 |
| Âge, 55-59 | -22,25 | -8,28 | 0,00 | 8,34 | 22,23 | -21,48 | 21,71 |
| Âge, 60-64 | -22,33 | -9,43 | 0,00 | 9,14 | 22,91 | -21,20 | 22,18 |
| Âge, 65-74 | -21,21 | -8,49 | 0,00 | 6,72 | 19,36 | -20,62 | 23,17 |
| Âge, > 74 | -26,23 | -11,99 | -1,31 | 7,74 | 22,02 | -24,90 | 23,60 |
| Célibataires | -7,26 | -2,34 | 0,00 | 2,33 | 7,14 | -13,63 | 13,11 |
| Personnes mariées | -6,33 | -1,99 | 0,00 | 2,31 | 6,98 | -8,50 | 8,72 |
| Veufs et veuves | -18,85 | -8,35 | -0,01 | 6,03 | 17,18 | -18,21 | 20,06 |
| Personnes divorcées | -19,62 | -7,28 | 0,00 | 7,72 | 19,43 | -21,57 | 20,79 |
| Personnes séparées | -20,50 | -7,85 | 0,00 | 8,55 | 22,51 | -22,90 | 20,28 |
| **Familles de recensement** | | | | | | | |
| Nombre total de familles | -4,42 | -1,52 | 0,00 | 1,54 | 4,69 | -7,15 | 7,48 |
| Familles époux-épouse | -4,91 | -1,69 | 0,00 | 1,77 | 5,48 | -8,19 | 8,18 |
| Familles monoparentales | -9,90 | -1,43 | 0,00 | 1,88 | 10,34 | -10,39 | 9,57 |
| Enfants dans les familles | -7,74 | -2,33 | 0,00 | 2,60 | 8,25 | -14,51 | 14,45 |
| Personnes dans les familles | -4,96 | -1,29 | 0,00 | 1,57 | 5,01 | -9,44 | 9,54 |
| Personnes hors famille | -16,76 | -6,51 | 0,00 | 4,59 | 14,28 | -19,36 | 19,04 |
| **Ménages et logements** | | | | | | | |
| Logements possédés | -4,97 | -1,61 | 0,00 | 1,73 | 4,84 | -6,74 | 6,85 |
| Logements loués | -11,45 | -2,88 | 0,00 | 2,74 | 9,52 | -13,97 | 13,60 |
| Maisons individuelles | -4,22 | -1,31 | 0,00 | 1,39 | 3,96 | -5,93 | 5,99 |
| App. dans un immeuble ≥ 5 étages | -3,52 | -0,80 | 0,00 | 0,55 | 4,24 | -6,43 | 7,26 |
| Logements mobiles | -11,81 | -4,43 | 0,00 | 3,15 | 9,29 | -12,54 | 15,80 |
| Autres logements | -7,35 | -1,69 | 0,00 | 2,71 | 9,81 | -11,77 | 11,77 |
| Nombre total de ménages | -1,55 | 0,00 | 0,00 | 0,00 | 1,39 | -4,45 | 4,50 |
| Ménages d'une personne | -11,93 | -5,07 | -0,54 | 3,51 | 10,46 | -15,74 | 15,47 |
| Ménages de 2 personnes | -11,13 | -3,71 | 0,00 | 4,02 | 11,56 | -15,78 | 16,38 |
| Ménages de 3 personnes | -15,33 | -5,54 | 0,00 | 6,18 | 17,21 | -20,29 | 20,59 |
| Ménages de 4 personnes | -14,78 | -5,03 | 0,00 | 4,44 | 13,50 | -19,22 | 18,31 |
| Ménages de 5 personnes | -14,04 | -4,94 | 0,40 | 8,11 | 19,45 | -21,86 | 22,06 |
| Ménages de 6 personnes et plus | -20,84 | -10,41 | -3,38 | 4,13 | 12,07 | -25,20 | 22,42 |
| Ménages non familiaux | -10,37 | -3,84 | -0,29 | 3,16 | 9,27 | -15,97 | 16,09 |
| Ménages unifamiliaux | -4,72 | -1,58 | 0,21 | 2,09 | 5,47 | -7,10 | 7,42 |
| Soutiens du ménage < 25 ans | -15,86 | -7,74 | 0,00 | 8,78 | 23,06 | -19,76 | 16,65 |
| Soutiens du ménage 25-34 ans | -14,42 | -4,73 | 0,00 | 4,43 | 13,73 | -16,94 | 15,25 |
| Soutiens du ménage 35-44 ans | -12,84 | -4,01 | 0,00 | 5,00 | 14,29 | -15,76 | 15,05 |
| Soutiens du ménage 45-54 ans | -15,87 | -4,83 | 0,00 | 5,98 | 17,57 | -16,63 | 15,60 |
| Soutiens du ménage 55-64 ans | -18,33 | -6,89 | 0,00 | 6,62 | 18,45 | -17,44 | 18,62 |
| Soutiens du ménage 65-74 ans | -18,01 | -6,98 | 0,00 | 5,84 | 16,95 | -17,97 | 19,44 |
| Soutiens du ménage > 74 ans | -20,81 | -9,52 | -0,42 | 7,08 | 17,22 | -21,07 | 20,86 |
| Soutiens du ménage sexe masculin | -5,23 | -1,65 | 0,00 | 1,75 | 5,24 | -7,63 | 7,22 |
| Soutiens du ménage sexe féminin | -10,58 | -3,02 | 0,00 | 3,06 | 10,82 | -15,84 | 18,12 |

**Tableau 10. Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SR – recensements de 1991 et de 1986**

| Caractéristiques étudiées | Centiles des écarts – 1991 | | | | | 1986 | |
|---|---|---|---|---|---|---|---|
| | 10e | 25e | 50e | 75e | 90e | 10e | 90e |
| **Personnes** | | | | | | | |
| Hommes | -0,88 | 0,00 | 0,00 | 0,00 | 0,87 | -2,13 | 1,94 |
| Femmes | -0,90 | 0,00 | 0,00 | 0,00 | 0,79 | -1,95 | 1,94 |
| Population totale | -0,24 | 0,00 | 0,00 | 0,00 | 0,24 | -1,67 | 1,53 |
| Âge, 0-4 | -10,77 | -1,16 | 0,00 | 1,03 | 9,77 | -6,50 | 6,83 |
| Âge, 5-9 | -10,97 | -1,55 | 0,00 | 0,94 | 10,16 | -6,80 | 6,68 |
| Âge, 10-14 | -11,19 | -1,34 | 0,00 | 2,23 | 13,00 | -6,80 | 6,48 |
| Âge, 15-19 | -10,92 | -2,72 | 0,00 | 3,93 | 11,62 | -6,86 | 7,03 |
| Âge, 20-24 | -10,09 | -1,48 | 0,00 | 1,90 | 10,82 | -6,79 | 6,92 |
| Âge, 25-29 | -9,97 | -1,57 | 0,00 | 0,28 | 7,31 | -12,80 | 12,15 |
| Âge, 30-34 | -7,78 | -0,57 | 0,00 | 0,29 | 7,46 | -12,59 | 12,74 |
| Âge, 35-39 | -9,25 | -1,32 | 0,00 | 0,00 | 7,88 | -13,00 | 12,62 |
| Âge, 40-44 | -10,00 | -1,43 | 0,00 | 1,02 | 9,05 | -14,42 | 14,56 |
| Âge, 45-49 | -12,65 | -3,06 | 0,00 | 3,06 | 12,00 | -15,39 | 15,01 |
| Âge, 50-54 | -14,93 | -5,00 | 0,00 | 5,84 | 16,01 | -15,88 | 15,84 |
| Âge, 55-59 | -16,32 | -4,96 | 0,00 | 5,84 | 17,34 | -16,05 | 16,05 |
| Âge, 60-64 | -17,42 | -6,53 | 0,00 | 6,54 | 18,61 | -16,96 | 17,34 |
| Âge, 65-74 | -14,16 | -3,00 | 0,00 | 1,61 | 12,41 | -12,49 | 12,89 |
| Âge, > 74 | -21,54 | -8,89 | 0,00 | 8,23 | 21,00 | -21,43 | 20,62 |
| Célibataires | -1,21 | 0,00 | 0,00 | 0,00 | 1,27 | -2,85 | 2,83 |
| Personnes mariées | -1,42 | 0,00 | 0,00 | 0,00 | 1,59 | -2,32 | 2,02 |
| Veufs et veuves | -15,41 | -5,27 | 0,00 | 3,87 | 14,87 | -16,21 | 17,02 |
| Personnes divorcées | -15,84 | -5,22 | 0,00 | 4,08 | 14,20 | -21,36 | 21,80 |
| Personnes séparées | -23,67 | -8,79 | 0,00 | 8,86 | 25,51 | -24,42 | 24,41 |
| **Familles de recensement** | | | | | | | |
| Nombre total de familles | -1,08 | 0,00 | 0,00 | 0,00 | 1,01 | -2,03 | 1,71 |
| Familles époux-épouse | -1,17 | 0,00 | 0,00 | 0,00 | 1,24 | -2,22 | 2,06 |
| Familles monoparentales | -6,03 | 0,00 | 0,00 | 0,00 | 5,40 | -5,68 | 6,15 |
| Enfants dans les familles | -1,77 | 0,00 | 0,00 | 0,26 | 2,12 | -3,28 | 3,40 |
| Personnes dans les familles | -0,91 | 0,00 | 0,00 | 0,08 | 1,02 | -2,07 | 1,88 |
| Personnes hors famille | -4,60 | -0,36 | 0,00 | 0,00 | 3,24 | -4,37 | 4,26 |
| **Ménages et logements** | | | | | | | |
| Logements possédés | -1,59 | 0,00 | 0,00 | 0,00 | 1,53 | -1,81 | 2,03 |
| Logements loués | -2,17 | 0,00 | 0,00 | 0,00 | 1,90 | -2,87 | 2,73 |
| Maisons individuelles | -1,22 | 0,00 | 0,00 | 0,11 | 1,32 | -2,90 | 2,97 |
| App. dans un immeuble ≥ 5 étages | -2,42 | 0,00 | 0,00 | 0,16 | 2,22 | -7,15 | 8,32 |
| Logements mobiles | -13,00 | -2,87 | 0,00 | 1,96 | 9,31 | -10,68 | 10,77 |
| Autres logements | -2,08 | -0,13 | 0,00 | 0,39 | 2,40 | -5,94 | 5,44 |
| Nombre total de ménages | -0,01 | 0,00 | 0,00 | 0,00 | 0,03 | -1,05 | 1,03 |
| Ménages d'une personne | -7,60 | -3,17 | -0,36 | 1,38 | 5,39 | -4,58 | 4,36 |
| Ménages de 2 personnes | -3,90 | -0,32 | 0,00 | 0,40 | 4,23 | -6,90 | 7,70 |
| Ménages de 3 personnes | -8,90 | -2,30 | 0,00 | 3,32 | 9,95 | -14,29 | 13,32 |
| Ménages de 4 personnes | -7,58 | -0,90 | 0,00 | 2,19 | 9,90 | -14,18 | 14,13 |
| Ménages de 5 personnes | -15,85 | -4,81 | 0,00 | 7,47 | 18,54 | -22,73 | 22,91 |
| Ménages de 6 personnes et plus | -23,65 | -11,04 | -0,54 | 7,11 | 16,34 | -31,21 | 26,78 |
| Ménages non familiaux | -5,70 | -2,20 | -0,29 | 1,29 | 4,08 | -4,34 | 3,95 |
| Ménages unifamiliaux | -2,09 | -0,79 | 0,24 | 1,23 | 2,43 | -2,10 | 2,08 |
| Soutiens du ménage < 25 ans | -19,94 | -8,97 | 0,00 | 8,95 | 20,02 | -22,02 | 20,37 |
| Soutiens du ménage 25-34 ans | -6,78 | -0,63 | 0,00 | 0,63 | 6,50 | -7,67 | 7,84 |
| Soutiens du ménage 35-44 ans | -5,31 | 0,00 | 0,00 | 0,00 | 5,33 | -8,45 | 8,15 |
| Soutiens du ménage 45-54 ans | -7,89 | -0,57 | 0,00 | 1,79 | 9,10 | -9,73 | 9,66 |
| Soutiens du ménage 55-64 ans | -10,30 | -1,92 | 0,00 | 2,44 | 11,77 | -10,65 | 11,33 |
| Soutiens du ménage 65-74 ans | -14,99 | -5,06 | 0,00 | 2,94 | 11,90 | -13,86 | 14,72 |
| Soutiens du ménage > 74 ans | -19,24 | -7,33 | 0,00 | 6,87 | 17,96 | -21,14 | 20,26 |
| Soutiens du ménage sexe masculin | -1,25 | 0,00 | 0,00 | 0,00 | 1,17 | -2,89 | 2,53 |
| Soutiens du ménage sexe féminin | -2,29 | 0,00 | 0,00 | 0,00 | 2,67 | -6,77 | 7,09 |

**Tableau 11. Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SRP – recensements de 1991 et de 1986**

| Caractéristiques étudiées | Centiles des écarts – 1991 | | | | | 1986 | |
|---|---|---|---|---|---|---|---|
| | 10e | 25e | 50e | 75e | 90e | 10e | 90e |
| **Personnes** | | | | | | | |
| Hommes | -1,48 | -0,58 | 0,00 | 0,54 | 1,33 | -2,55 | 2,68 |
| Femmes | -1,44 | -0,56 | 0,00 | 0,59 | 1,54 | -2,71 | 2,60 |
| Population totale | -0,84 | -0,26 | 0,00 | 0,29 | 0,87 | -2,11 | 2,06 |
| Âge, 0-4 | -10,98 | -4,99 | 0,00 | 4,02 | 10,23 | -8,56 | 8,72 |
| Âge, 5-9 | -9,57 | -4,01 | 0,00 | 3,90 | 9,46 | -8,47 | 8,82 |
| Âge, 10-14 | -9,53 | -3,86 | 0,00 | 4,22 | 9,91 | -8,59 | 8,60 |
| Âge, 15-19 | -8,88 | -3,78 | 0,00 | 5,00 | 11,54 | -8,64 | 8,78 |
| Âge, 20-24 | -13,17 | -5,47 | 0,00 | 5,75 | 14,26 | -9,80 | 9,41 |
| Âge, 25-29 | -11,71 | -4,77 | 0,00 | 3,41 | 9,43 | -12,54 | 12,62 |
| Âge, 30-34 | -9,22 | -3,35 | 0,00 | 3,69 | 9,88 | -12,24 | 12,01 |
| Âge, 35-39 | -10,28 | -4,22 | 0,00 | 3,75 | 9,92 | -12,40 | 13,11 |
| Âge, 40-44 | -10,20 | -4,36 | 0,00 | 4,64 | 11,19 | -13,88 | 14,99 |
| Âge, 45-49 | -12,13 | -5,43 | 0,00 | 6,27 | 13,76 | -15,71 | 15,43 |
| Âge, 50-54 | -16,36 | -7,16 | 0,00 | 7,58 | 16,29 | -15,90 | 16,13 |
| Âge, 55-59 | -16,66 | -7,56 | 0,00 | 6,82 | 15,40 | -16,24 | 16,60 |
| Âge, 60-64 | -16,11 | -7,68 | 0,00 | 7,49 | 16,88 | -16,40 | 16,40 |
| Âge, 65-74 | -11,33 | -4,33 | 0,00 | 3,82 | 10,00 | -11,13 | 12,57 |
| Âge, > 74 | -21,49 | -10,26 | -1,17 | 6,13 | 16,49 | -21,08 | 17,98 |
| Célibataires | -1,90 | -0,68 | 0,00 | 0,80 | 1,98 | -3,86 | 3,91 |
| Personnes mariées | -1,67 | -0,68 | 0,00 | 0,83 | 1,95 | -2,51 | 2,41 |
| Veufs et veuves | -14,28 | -6,34 | 0,00 | 3,92 | 10,82 | -14,31 | 14,77 |
| Personnes divorcées | -16,70 | -7,78 | 0,00 | 6,80 | 16,68 | -22,77 | 25,18 |
| Personnes séparées | -23,27 | -10,35 | 0,00 | 11,45 | 25,79 | -25,85 | 26,49 |
| **Familles de recensement** | | | | | | | |
| Nombre total de familles | -1,29 | -0,52 | 0,00 | 0,55 | 1,25 | -2,06 | 2,19 |
| Familles époux-épouse | -1,45 | -0,61 | 0,00 | 0,63 | 1,55 | -2,40 | 2,43 |
| Familles monoparentales | -10,56 | -3,92 | 0,00 | 3,73 | 10,25 | -11,62 | 10,83 |
| Enfants dans les familles | -2,18 | -0,79 | 0,00 | 1,08 | 2,64 | -4,25 | 4,25 |
| Personnes dans les familles | -1,26 | -0,44 | 0,00 | 0,58 | 1,32 | -2,47 | 2,59 |
| Personnes hors famille | -6,38 | -2,68 | 0,00 | 1,74 | 4,89 | -7,20 | 7,11 |
| **Ménages et logements** | | | | | | | |
| Logements possédés | -1,53 | -0,62 | 0,00 | 0,64 | 1,65 | -2,18 | 2,08 |
| Logements loués | -6,03 | -1,77 | 0,00 | 1,63 | 5,32 | -6,58 | 6,40 |
| Maisons individuelles | -1,36 | -0,49 | 0,00 | 0,53 | 1,31 | -2,35 | 2,45 |
| App. dans un immeuble ≥ 5 étages | -6,05 | -0,60 | 0,00 | 0,76 | 2,22 | -11,44 | 9,99 |
| Logements mobiles | -14,68 | -5,67 | 0,00 | 4,85 | 12,50 | -17,31 | 17,11 |
| Autres logements | -6,41 | -1,74 | 0,00 | 2,41 | 6,90 | -9,60 | 9,03 |
| Nombre total de ménages | -0,54 | -0,11 | 0,00 | 0,14 | 0,56 | -1,34 | 1,38 |
| Ménages d'une personne | -7,11 | -3,60 | -0,54 | 2,17 | 6,28 | -8,37 | 8,05 |
| Ménages de 2 personnes | -4,65 | -1,77 | 0,00 | 1,88 | 4,76 | -7,94 | 8,03 |
| Ménages de 3 personnes | -10,00 | -4,39 | 0,00 | 4,65 | 10,27 | -14,18 | 14,42 |
| Ménages de 4 personnes | -8,76 | -3,57 | 0,00 | 3,81 | 8,51 | -11,95 | 13,06 |
| Ménages de 5 personnes | -14,77 | -5,82 | 0,88 | 9,24 | 18,62 | -20,93 | 19,79 |
| Ménages de 6 personnes et plus | -27,14 | -15,35 | -4,22 | 7,39 | 18,29 | -27,92 | 27,64 |
| Ménages non familiaux | -5,24 | -2,50 | -0,30 | 1,73 | 4,34 | -8,14 | 7,08 |
| Ménages unifamiliaux | -1,70 | -0,63 | 0,20 | 1,07 | 2,00 | -2,20 | 2,30 |
| Soutiens du ménage < 25 ans | -23,55 | -11,01 | -0,50 | 11,66 | 25,39 | -22,44 | 20,89 |
| Soutiens du ménage 25-34 ans | -7,61 | -3,37 | 0,00 | 3,03 | 7,97 | -8,14 | 8,53 |
| Soutiens du ménage 35-44 ans | -6,44 | -2,34 | 0,00 | 2,93 | 7,26 | -8,76 | 9,27 |
| Soutiens du ménage 45-54 ans | -8,78 | -3,83 | 0,00 | 4,41 | 10,85 | -10,72 | 10,96 |
| Soutiens du ménage 55-64 ans | -11,71 | -5,00 | 0,00 | 5,04 | 11,28 | -11,00 | 11,08 |
| Soutiens du ménage 65-74 ans | -12,21 | -5,13 | 0,00 | 4,15 | 11,59 | -13,29 | 15,06 |
| Soutiens du ménage > 74 ans | -21,70 | -10,26 | -0,87 | 6,59 | 16,36 | -22,91 | 18,59 |
| Soutiens du ménage sexe masculin | -1,48 | -0,65 | 0,00 | 0,59 | 1,49 | -2,77 | 2,55 |
| Soutiens du ménage sexe féminin | -4,99 | -1,77 | 0,00 | 2,07 | 4,95 | -8,33 | 9,09 |

## D. Secteurs de dénombrement (SD)

Les SD sont les composantes des RP, et les RP sont le niveau le plus bas auquel les estimations-échantillon sont forcées de correspondre aux chiffres de population pour la plupart des caractéristiques. Les SD sont aussi les composantes de niveaux géographiques plus élevés (DR, SDR, SR, SRP, etc.) et plusieurs des RP, comme le montrait le tableau 3, sont des composantes de ces niveaux supérieurs. Par conséquent, on ne peut pas s'attendre à ce que la cohérence au niveau des SD soit aussi bonne qu'aux niveaux géographiques plus élevés qui ont été étudiés. Le tableau 12 le confirme : pour la plupart des caractéristiques étudiées, dans les SD échantillonnés ayant un chiffre de population pour la caractéristique supérieur à 50, les écarts sont supérieurs à ceux observés aux niveaux géographiques examinés précédemment. C'est le cas aussi bien en 1991 qu'en 1986. Par rapport aux écarts de 1986 au 10e et au 90e centile, les écarts de 1991 sont beaucoup plus faibles pour l'immense majorité des caractéristiques étudiées, et semblables à ceux de 1986 pour les quelques caractéristiques qui restent.

Une étude semblable a été faite précédemment pour les mêmes niveaux géographiques. Un total de 68 caractéristiques, incluant les 53 étudiées ci-dessus, ont été examinées, et l'on a obtenu les écarts tant pour les estimations de 1991 que pour celles de 1986 aux 10e, 25e, 50e, 75e et 90e centiles. Pour plus de renseignements sur cette étude, voir Majkowski (1992a).

**Tableau 12. Centiles des écarts entre les estimations-échantillon et les chiffres de population (en pourcentage des chiffres de population) pour les SD – recensements de 1991 et de 1986**

| Caractéristiques étudiées | Centiles des écarts – 1991 | | | | | 1986 | |
|---|---|---|---|---|---|---|---|
| | 10e | 25e | 50e | 75e | 90e | 10e | 90e |
| **Personnes** | | | | | | | |
| Hommes | -6,25 | -2,59 | 0,07 | 2,64 | 6,06 | -12,81 | 13,01 |
| Femmes | -6,29 | -2,66 | -0,08 | 2,55 | 6,24 | -12,01 | 12,16 |
| Population totale | -3,56 | 0,00 | 0,00 | 0,00 | 3,35 | -9,83 | 9,99 |
| Âge, 0-4 | -26,37 | -12,73 | -0,22 | 12,50 | 26,49 | -29,25 | 30,17 |
| Âge, 5-9 | -24,17 | -11,49 | 0,03 | 11,23 | 24,00 | -28,54 | 28,75 |
| Âge, 10-14 | -25,42 | -12,24 | 0,07 | 12,23 | 25,28 | -28,48 | 28,87 |
| Âge, 15-19 | -25,38 | -12,06 | 0,37 | 13,15 | 27,12 | -29,55 | 30,57 |
| Âge, 20-24 | -27,94 | -13,58 | -0,23 | 13,51 | 28,48 | -29,57 | 30,39 |
| Âge, 25-29 | -26,08 | -12,60 | -0,51 | 11,45 | 25,23 | -30,21 | 29,88 |
| Âge, 30-34 | -24,43 | -11,60 | -0,23 | 11,43 | 24,63 | -30,13 | 30,59 |
| Âge, 35-39 | -25,09 | -12,36 | -0,68 | 11,28 | 24,60 | -29,83 | 30,18 |
| Âge, 40-44 | -26,31 | -12,80 | -0,37 | 11,93 | 26,53 | -31,06 | 31,66 |
| Âge, 45-49 | -27,42 | -13,12 | -0,02 | 13,65 | 29,16 | -31,97 | 32,29 |
| Âge, 50-54 | -30,00 | -15,15 | -0,10 | 14,85 | 31,77 | -31,25 | 33,71 |
| Âge, 55-59 | -29,56 | -15,01 | 0,29 | 15,76 | 31,40 | -31,59 | 34,43 |
| Âge, 60-64 | -30,34 | -15,37 | -0,50 | 16,01 | 31,98 | -32,51 | 35,26 |
| Âge, 65-74 | -26,62 | -12,59 | -0,63 | 11,72 | 27,41 | -30,08 | 31,33 |
| Âge, > 74 | -27,47 | -12,72 | -0,32 | 12,50 | 27,83 | -30,54 | 32,05 |
| Célibataires | -8,45 | -3,43 | 0,04 | 3,42 | 8,03 | -17,47 | 17,67 |
| Personnes mariées | -8,78 | -3,55 | 0,07 | 3,76 | 8,92 | -13,10 | 12,95 |
| Veufs et veuves | -23,51 | -11,54 | -0,60 | 10,68 | 22,99 | -26,94 | 29,00 |
| Personnes divorcées | -29,67 | -15,22 | -0,93 | 14,50 | 29,84 | -32,41 | 35,25 |
| Personnes séparées | -28,90 | -11,74 | 3,21 | 17,38 | 32,19 | -33,44 | 36,57 |
| **Familles de recensement** | | | | | | | |
| Nombre total de familles | -5,72 | -2,53 | -0,02 | 2,48 | 5,80 | -10,62 | 10,61 |
| Familles époux-épouse | -6,54 | -2,95 | 0,01 | 2,93 | 6,60 | -12,31 | 12,16 |
| Familles monoparentales | -17,57 | -8,93 | 0,61 | 10,14 | 19,57 | -26,38 | 24,83 |
| Enfants dans les familles | -9,98 | -3,88 | 0,19 | 4,13 | 9,81 | -19,94 | 20,38 |
| Personnes dans les familles | -6,39 | -2,31 | 0,08 | 2,37 | 6,31 | -13,30 | 13,31 |
| Personnes hors famille | -19,17 | -8,92 | -0,25 | 7,96 | 18,09 | -24,31 | 25,02 |
| **Ménages et logements** | | | | | | | |
| Logements possédés | -7,16 | -3,01 | 0,09 | 3,02 | 7,05 | -10,15 | 10,13 |
| Logements loués | -9,94 | -3,82 | 0,11 | 4,00 | 9,77 | -13,82 | 13,32 |
| Maisons individuelles | -6,17 | -2,51 | 0,02 | 2,48 | 6,13 | -8,73 | 8,87 |
| App. dans un immeuble ≥ 5 étages | -6,37 | -2,23 | 0,13 | 2,39 | 6,40 | -9,60 | 9,98 |
| Logements mobiles | -12,40 | -4,74 | 0,10 | 5,14 | 11,19 | -14,69 | 16,40 |
| Autres logements | -8,44 | -3,25 | 0,08 | 3,43 | 8,48 | -12,44 | 11,88 |
| Nombre total de ménages | -0,18 | 0,00 | 0,00 | 0,00 | 0,04 | -6,31 | 6,23 |
| Ménages d'une personne | -13,75 | -6,67 | -0,35 | 5,69 | 12,38 | -21,47 | 21,42 |
| Ménages de 2 personnes | -14,41 | -7,08 | -0,04 | 7,06 | 15,11 | -22,07 | 23,55 |
| Ménages de 3 personnes | -22,27 | -11,37 | -0,12 | 11,25 | 23,48 | -27,75 | 28,63 |
| Ménages de 4 personnes | -17,59 | -8,41 | -0,04 | 8,49 | 18,45 | -24,95 | 26,08 |
| Ménages de 5 personnes | -22,17 | -10,99 | 1,09 | 13,92 | 26,33 | -30,33 | 31,62 |
| Ménages de 6 personnes et plus | -22,63 | -9,30 | 2,48 | 14,81 | 26,63 | -38,54 | 27,49 |
| Ménages non familiaux | -11,54 | -5,28 | -0,04 | 4,91 | 11,03 | -20,18 | 20,25 |
| Ménages unifamiliaux | -6,78 | -2,95 | 0,34 | 3,44 | 7,01 | -10,82 | 11,06 |
| Soutiens du ménage < 25 ans | -21,82 | -10,84 | 0,95 | 12,93 | 24,53 | -28,01 | 31,55 |
| Soutiens du ménage 25-34 ans | -17,55 | -8,40 | 0,06 | 8,43 | 18,27 | -23,63 | 24,19 |
| Soutiens du ménage 35-44 ans | -18,12 | -8,66 | -0,11 | 8,34 | 18,77 | -23,81 | 24,39 |
| Soutiens du ménage 45-54 ans | -21,43 | -10,20 | -0,07 | 10,61 | 22,95 | -25,43 | 26,29 |
| Soutiens du ménage 55-64 ans | -23,30 | -11,39 | -0,08 | 11,35 | 24,20 | -26,66 | 27,51 |
| Soutiens du ménage 65-74 ans | -19,82 | -9,90 | -0,51 | 9,19 | 19,91 | -26,35 | 28,44 |
| Soutiens du ménage > 74 ans | -18,88 | -8,90 | -0,08 | 9,68 | 19,34 | -25,04 | 27,55 |
| Soutiens du ménage sexe masculin | -6,88 | -3,19 | 0,05 | 3,25 | 6,84 | -11,67 | 11,46 |
| Soutiens du ménage sexe féminin | -12,61 | -6,19 | -0,09 | 6,10 | 13,10 | -21,31 | 21,89 |

# VIII. Variance d'échantillonnage

L'erreur d'échantillonnage a deux composantes : la variance et le biais. La variance mesure la variabilité de l'estimation autour de l'estimation moyenne qui résulterait de répétitions hypothétiques du processus d'enquête. Le biais est la différence entre la valeur moyenne d'une estimation qui résulterait de répétitions hypothétiques du processus d'enquête, et la valeur vraie de la variable estimée. L'erreur quadratique moyenne (EQM) mesure la variabilité de l'estimation autour de la valeur vraie, pour des répétitions hypothétiques du processus d'enquête. On peut montrer que l'EQM est égale à la somme de la variance et du carré du biais. L'EQM est la mesure la plus exacte de l'ampleur moyenne de l'écart entre l'estimation et la valeur vraie de la population. Si le biais est faible par rapport à la variance, la variance est une bonne approximation de l'EQM. On constate toutefois qu'en réalité, le biais s'accumule à mesure qu'on produit des estimations du recensement visant des unités géographiques plus grandes. Le biais peut donc être non significatif pour de petites unités géographiques, mais devenir élevé par rapport à la variance pour de grandes unités géographiques. Il est possible, par conséquent, que la variance soit beaucoup plus faible que l'EQM pour de grandes unités géographiques. On peut estimer la variance d'une estimation à partir de l'échantillon, mais on ne peut pas faire de même dans le cas du biais d'une estimation. Autrement dit, il n'est pas possible d'estimer avec exactitude l'EQM à partir de l'échantillon, à moins que le biais soit faible par rapport à la variance.

Dans des recensements précédents, une étude visant à produire des estimations de la variance d'échantillonnage a été réalisée. La section A donne quelques résultats de l'étude de 1986 (pour plus de renseignements, voir le Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986 : Échantillonnage et pondération). Toutefois, parce qu'on croyait que la variance d'échantillonnage ne donnerait pas une estimation exacte de l'EQM pour les grandes unités géographiques, il a été décidé de ne pas reprendre cette étude pour le recensement de 1991. On trouvera à la section B, cependant, quelques observations relatives à l'impact qu'a eu la méthode d'estimation utilisée dans le recensement de 1991 sur la variance d'échantillonnage, par rapport au recensement de 1986.

## A. Étude sur la variance d'échantillonnage du recensement de 1986

Le chapitre V, qui contient les résultats de l'examen du biais d'échantillonnage, décrit la nature et l'importance du biais dans l'échantillon avant la pondération. Les chapitres VI et VII examinent le biais d'échantillonnage après la pondération. Même avec une méthode d'échantillonnage totalement dépourvue de biais, les résultats seraient sujets à la variance du simple fait que les estimations ont été produites à partir d'un échantillon. La variance peut être estimée à l'aide des données recueillies dans l'enquête-échantillon[12]. L'étude de la variance d'échantillonnage de 1986 a été effectuée afin d'estimer l'effet des méthodes d'échantillonnage et d'estimation sur les chiffres du recensement établis à partir de données-échantillon.

Statistique Canada produit des milliers de tableaux à partir des données-échantillon (c'est-à-dire celles recueillies sur la formule 2B). Théoriquement, on pourrait fournir une mesure de précision (à savoir l'estimation de la variance d'échantillonnage) pour chaque estimation présentée dans ces tableaux. Cette mesure tiendrait compte à la fois du plan de sondage et de la méthode d'estimation. En pratique cependant, il est impossible de calculer une telle mesure pour toutes les estimations du recensement en raison du coût élevé du traitement des données. On calcule donc la variance d'échantillonnage pour un sous-ensemble seulement des estimations du recensement. Cela permet ensuite d'évaluer les effets combinés du plan de sondage et de la méthode d'estimation sur la variance d'échantillonnage. On peut ensuite ajuster les estimations simples de la variance d'échantillonnage, qui sont peu coûteuses à produire, pour tenir compte de cet effet et produire une estimation de la variance d'échantillonnage pour n'importe quelle estimation du recensement.

Le tableau 13 donne des valeurs non ajustées (simples) de l'erreur-type des estimations-échantillon du recensement. Les valeurs présentées dans ce tableau ont été établies en supposant qu'on a utilisé un échantillon aléatoire simple avec fraction de sondage de 1 sur 5, et une pondération simple de 5. Les erreurs-types sont présentées au tableau 13

12 Malheureusement, la variance d'échantillonnage ne nous fournit aucune indication de l'importance de l'erreur non due à l'échantillonnage.

en fonction à la fois de l'ampleur du «total estimé» et de la taille de l'unité. Par exemple, pour une estimation de 250 personnes dans une unité géographique comptant 1 000 personnes, l'erreur-type non ajustée est de 25.

**Tableau 13. Estimations des erreurs-types des estimations-échantillon**

| Total estimé | Nombre total de personnes, ménages, logements ou familles dans l'unité | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 500 | 1 000 | 2 500 | 5 000 | 10 000 | 25 000 | 50 000 | 100 000 | 250 000 |
| 50 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 |
| 100 | 18 | 19 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| 250 | 22 | 25 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| 500 | 0 | 30 | 40 | 40 | 45 | 45 | 45 | 45 | 45 |
| 1 000 | | 0 | 50 | 55 | 60 | 60 | 65 | 65 | 65 |
| 2 500 | | | 0 | 70 | 85 | 95 | 95 | 100 | 100 |
| 5 000 | | | | 0 | 100 | 130 | 130 | 140 | 140 |
| 10 000 | | | | | 0 | 150 | 180 | 190 | 200 |
| 25 000 | | | | | | 0 | 220 | 270 | 300 |
| 50 000 | | | | | | | 0 | 320 | 400 |
| 100 000 | | | | | | | | 0 | 490 |
| 250 000 | | | | | | | | | 0 |

| Total estimé | Nombre total de personnes, ménages, logements ou familles dans l'unité | | | | | |
|---|---|---|---|---|---|---|
| | 500 000 | 1 000 000 | 2 500 000 | 5 000 000 | 10 000 000 | 25 000 000 |
| 50 | 15 | 15 | 15 | 15 | 15 | 15 |
| 100 | 20 | 20 | 20 | 20 | 20 | 20 |
| 250 | 30 | 30 | 30 | 30 | 30 | 30 |
| 500 | 45 | 45 | 45 | 45 | 45 | 45 |
| 1 000 | 65 | 65 | 65 | 65 | 65 | 65 |
| 2 500 | 100 | 100 | 100 | 100 | 100 | 100 |
| 5 000 | 140 | 140 | 140 | 140 | 140 | 140 |
| 10 000 | 200 | 200 | 200 | 200 | 200 | 200 |
| 25 000 | 310 | 310 | 310 | 320 | 320 | 320 |
| 50 000 | 420 | 440 | 440 | 440 | 450 | 450 |
| 100 000 | 570 | 600 | 620 | 630 | 630 | 630 |
| 250 000 | 710 | 870 | 950 | 970 | 990 | 990 |
| 500 000 | 0 | 1 000 | 1 260 | 1 340 | 1 380 | 1 400 |
| 1 000 000 | | 0 | 1 550 | 1 790 | 1 900 | 1 960 |
| 2 500 000 | | | 0 | 2 240 | 2 740 | 3 000 |
| 5 000 000 | | | | 0 | 3 160 | 4 000 |
| 10 000 000 | | | | | 0 | 4 900 |

Le tableau 13 contient les erreurs-types associées à un nombre restreint de valeurs du «total estimé» et du nombre total de personnes, de ménages, de logements ou de familles dans l'unité. La formule qui suit peut servir à calculer l'erreur-type non ajustée de n'importe quel total estimé, pour une unité de n'importe quelle taille :

$$\text{ETNA} = \sqrt{\frac{4E(N - E)}{N}} \qquad (4)$$

où ETNA est l'erreur-type non ajustée, E le total estimé et N le nombre total de personnes, de ménages, de logements ou de familles dans l'unité. Par exemple, pour un total estimé de 750 personnes dans une unité comptant 9 000 personnes, l'erreur-type non ajustée serait :

$$\sqrt{\frac{4(750)(9000 - 750)}{9000}} = 52$$

Il convient de signaler que si E est beaucoup plus petit que N, les valeurs N-E et N seront approximativement égales et ETNA, dans l'équation (4), sera alors approximativement égale à deux fois la racine carrée de l'estimation elle-même ($\text{ETNA} \simeq 2\sqrt{E}$).

L'étude sur la variance d'échantillonnage de 1986 fournit des facteurs d'ajustement[13] par lesquels on doit multiplier les erreurs-types non ajustées pour tenir compte des effets combinés du plan de sondage et de la méthode d'estimation. Pour calculer ces facteurs d'ajustement, un échantillon de 401 RP (sur un total de 5 341) a été tiré. On a réparti l'échantillon parmi les 10 provinces[14] de manière à obtenir de bonnes estimations de la variance d'échantillonnage au niveau provincial sans pour autant compromettre indûment la qualité des estimations au niveau national. Pour chaque RP dans l'échantillon, les estimations de la variance d'échantillonnage des estimations produites par la MIQ ont été calculées pour différentes sous-catégories de toutes les caractéristiques énumérées au tableau 9 du Guide à l'intention des utilisateurs du recensement de 1986. Les estimations de la variance d'échantillonnage aux niveaux provincial et national ont été obtenues en pondérant les estimations au niveau de la RP. Les facteurs d'ajustement pour chaque sous-catégorie de chacune des caractéristiques ont été calculés en divisant la racine carrée de ces estimations par l'erreur-type non ajustée correspondante. Des facteurs d'ajustement ont été calculés aux niveaux provincial et national pour chaque caractéristique en faisant la moyenne des facteurs d'ajustement de toutes les sous-catégories. Pour plus de renseignements sur la façon dont ces facteurs d'ajustement ont été calculés, voir Béland (1990).

Pour estimer l'erreur-type d'une estimation-échantillon donnée, le facteur d'ajustement s'appliquant à la caractéristique en question a été déterminé d'après le tableau 9 du Guide à l'intention des utilisateurs du recensement de 1986. Le facteur d'ajustement au niveau national ou provincial pour les caractéristiques-échantillon se situait généralement entre 0,40 et 1,60. Ce facteur a ensuite été multiplié par l'erreur-type non ajustée correspondante dans le tableau 13.

L'exemple qui suit illustre la façon de calculer l'erreur-type ajustée. Supposons que l'estimation étudiée soit la population immigrante de l'Ontario. L'estimation du recensement de 1986 pour cette caractéristique était 2 081 200. Le chiffre de population obtenu pour l'Ontario au recensement de 1986 était 9 001 170. Comme aucun des deux chiffres n'est très proche des valeurs données dans le tableau 13, on applique l'équation (4) pour obtenir l'erreur-type non ajustée. Dans ce cas-ci, le résultat serait 2 530. D'après le tableau 9 du Guide à l'intention des utilisateurs du recensement de 1986, le facteur d'ajustement au niveau provincial pour la caractéristique «immigrant» est 1,12. Il s'ensuit que l'erreur-type ajustée pour cette estimation est 2 530 x 1,12 = 2 834.

Un deuxième exemple, toutefois, soulève un doute quant à l'exactitude de ces erreurs-types ajustées à titre d'estimations de la racine carrée de l'EQM. L'estimation, dans le recensement de 1986, du nombre de personnes mariées vivant dans des logements privés dans les SD échantillonnés était 11 771 126. Le nombre de personnes, lors du recensement de 1986, qui vivaient dans des logements privés dans des SD échantillonnés était 24 369 559. L'application de l'équation (4) donne une erreur-type non ajustée de 4 934. D'après le tableau 9 du Guide à

13 On appelle communément «effet du plan» le carré des facteurs d'ajustement.

14 Le Yukon et les Territoires du Nord-Ouest ont été groupés avec la Colombie-Britannique.

l'intention des utilisateurs du recensement de 1986, le facteur d'ajustement au niveau national pour la caractéristique «marié» est 0,25. Par conséquent, l'erreur-type ajustée pour cette estimation est 4 934 x 0,25 = 1 233. Toutefois, puisque l'état matrimonial est une caractéristique de base, on sait que le chiffre de population relatif au nombre de personnes mariées lors du recensement de 1986 qui vivaient dans des logements privés dans des SD échantillonnés était 11 778 842. La différence entre l'estimation et le chiffre de population est -7 716. Le quotient entre cette différence et l'erreur-type ajustée est -7 716/1 233 = -6,26. Un intervalle de confiance à 95 % pour une estimation serait normalement défini comme étant plus ou moins deux fois l'erreur-type ajustée. Le fait que le quotient entre la différence et l'erreur-type soit de -6,26 laisse entendre que l'erreur-type ajustée de 1 233 est une sous-estimation de la racine carrée de l'EQM.

## B. Variance d'échantillonnage et biais dans le recensement de 1991

Dans Bankier, Rathwell et Majkowski (1992), les coefficients de variation (CV) de la PEGMC pour certaines caractéristiques-échantillon ont été comparés aux CV correspondants de la MIQ. Dans les deux cas, les données du recensement de 1986 ont été utilisées. Le CV d'une estimation est la racine carrée de la variance estimée exprimée en pourcentage de l'estimation. Pour 79 RP, on a estimé les CV pour les estimations de 507 caractéristiques-échantillon au niveau des SD et 642 caractéristiques-échantillon au niveau des RP (chacune d'entre elles s'appliquant à un total minimum estimé de 60 ménages dans la population). Chacune des estimations au niveau des RP et au niveau des SD a été classée comme estimation faible (inférieure ou égale à la médiane des estimations) ou comme estimation élevée (supérieure à la médiane des estimations). On a constaté que la valeur médiane des CV pour les estimations élevées au niveau des RP était de 5 % pour la PEGMC, tandis qu'elle était de 6 % pour la MIQ. La valeur médiane des CV pour les estimations faibles au niveau des RP était de 13 % pour la PEGMC et de 15 % pour la MIQ. La valeur médiane des CV pour les estimations élevées au niveau des SD était de 10 % pour la PEGMC, tandis qu'elle était de 12,5 % pour la MIQ. La valeur médiane des CV pour les estimations faibles au niveau des SD était de 15 % pour la PEGMC et de 17,5 % pour la MIQ. La PEGMC a donc permis une réduction des CV par rapport à la MIQ tant au niveau des SD qu'à celui des RP. Puisque les variances à des niveaux géographiques supérieurs sont simplement la somme des variances au niveau des RP, ces réductions des CV devraient se répercuter sur les niveaux géographiques plus élevés.

Il a été indiqué au chapitre V que l'échantillon du recensement affichait des biais faibles, mais significatifs. Ces biais ne sont pas significatifs par rapport à la variance d'échantillonnage au niveau des RP. À des niveaux géographiques supérieurs, toutefois, le biais touchant une caractéristique peut s'accumuler s'il produit systématiquement des surestimations ou des sous-estimations. Il semble que l'effet du biais soit plus important dans le cas de la PEGMC que dans celui de la MIQ. On peut le voir au tableau 7 du chapitre VI, où la PEGMC présente des différences population/estimation plus faibles que la MIQ pour les plus petites unités géographiques. La situation est inversée, toutefois, dans le cas des grandes unités géographiques. Outre le biais introduit par l'échantillonnage et le traitement, Bankier, Rathwell et Majkowski (1992) montrent, par une étude de Monte Carlo, que l'estimateur PEGMC est lui-même biaisé, bien que le biais relatif soit inférieur à 1 % pour 50 % des caractéristiques étudiées. Fait plus important, toutefois, la variance estimée des estimateurs PEGMC affiche un biais relatif médian de -25 % au niveau des RP. Ces estimateurs ont donc tendance à sous-estimer la variance vraie. Les estimateurs MIQ souffrent peut-être de biais semblables, mais aucune étude n'a été faite à ce sujet.

Pour le recensement de 1996, des améliorations seront apportées aux méthodes d'estimation de manière à réduire l'ampleur des différences population/estimation aux niveaux géographiques supérieurs. On devrait ainsi obtenir des estimations plus exactes des EQM des caractéristiques-échantillon.

# IX. Conclusion

L'échantillonnage fait maintenant partie intégrante du processus de recensement. Son utilisation permet de faire des économies substantielles et de réduire de façon appréciable le fardeau du répondant, ou encore, d'élargir la portée d'un recensement sans frais supplémentaires. Le prix de ces avantages est l'introduction d'une erreur d'échantillonnage dans les estimations obtenues à partir de l'échantillon. Cet effet se fait particulièrement sentir sur les estimations de faible valeur du recensement, qu'il s'agisse des chiffres se rapportant à des catégories peu usuelles au niveau national ou provincial, ou de ceux qui correspondent à des unités géographiques de petite taille. Il convient de souligner que les erreurs de réponse et les erreurs de dépouillement contribuent également à l'erreur totale des estimations du recensement et, là encore, ce sont les estimations de faible valeur qui sont particulièrement sensibles aux effets de ces erreurs non dues à l'échantillonnage. Par conséquent, bon nombre de ces faibles valeurs n'auraient qu'une fiabilité relative même si elles étaient fondées sur un recensement intégral. Pour ce qui est du recensement de 1991, les estimations inférieures ou égales à 50 établies à partir des données-échantillon sont généralement très peu fiables tandis que les estimations allant jusqu'à 500 tendent à avoir des erreurs-types de plus de 10 % de leur valeur.

Pour de nombreuses caractéristiques, un certain biais a été détecté dans l'échantillon. Il a été établi que ce biais avait été causé partiellement par le traitement des données et le Contrôle et Imputation. Le reste du biais provient sans doute d'un ou plusieurs facteurs tels que le biais de non-réponse, le biais de réponse, la sélection par les recenseurs d'un échantillon biaisé. Les méthodes utilisées pour pondérer les données-échantillon en vue d'obtenir des estimations à l'échelle de la population totale ont été appliquées avec succès et, de façon générale, ont permis d'atteindre les niveaux prévus de cohérence des estimations-échantillon et des chiffres de population. La cohérence obtenue aux niveaux provincial et national a été un peu plus faible que prévu, compte tenu de la cohérence améliorée atteinte lors des tests pour les plus petits niveaux géographiques. Cela s'explique probablement par le biais que comportait l'échantillon, ainsi que par l'introduction d'un léger biais additionnel attribuable à la méthode d'estimation elle-même.

La méthode d'estimation du recensement sera réévaluée en vue du recensement de 1996, afin de voir s'il est possible d'améliorer la cohérence entre les estimations-échantillon et les chiffres de population au niveau des provinces et du Canada, tout en maintenant une bonne cohérence au niveau des SD. On devrait pouvoir obtenir en même temps des estimations plus fiables de l'erreur quadratique moyenne des estimations du recensement.

# Annexe A

## Contraintes au niveau des RP et des SD appliquées aux coefficients de pondération du recensement de 1991 désignations des contraintes (formule abrégée)

### Contraintes au niveau des RP (personnes)

| | |
|---|---|
| TOTPERS | –Population totale |
| TPERGE15 | –Population totale d'âge ≥ 15 |
| MALE | –Hommes |
| MALEGE15 | –Hommes ≥ 15 ans |
| AGE4 | –Âge, 0 à 4 |
| AGE9 | –Âge, 5 à 9 |
| AGE14 | –Âge, 10 à 14 |
| AGE19 | –Âge, 15 à 19 |
| AGE24 | –Âge, 20 à 24 |
| AGE29 | –Âge, 25 à 29 |
| AGE34 | –Âge, 30 à 34 |
| AGE39 | –Âge, 35 à 39 |
| AGE44 | –Âge, 40 à 44 |
| AGE49 | –Âge, 45 à 49 |
| AGE54 | –Âge, 50 à 54 |
| AGE59 | –Âge, 55 à 59 |
| AGE64 | –Âge, 60 à 64 |
| AGE74 | –Âge, 65 à 74 |
| AGE75P | –Âge, ≥ 75 |
| MARRIED | –Personnes mariées |
| SINGLE | –Célibataires |
| DIVORCED | –Personnes divorcées |
| WIDOWED | –Veufs et veuves |
| SEP | –Personnes séparées |
| CENFAM | –Familles de recensement |
| NONMEMB | –Personnes hors famille de recensement |
| HUSBAND | –Époux |
| CHILD | –Enfants dans les familles de recensement |
| LONEPARF | –Mères seules |

### Contraintes au niveau des SD

| | |
|---|---|
| HHEACT | –Total des ménages dans le SD |
| PPEACT | –Population totale du SD |

### Contraintes au niveau des RP (ménages)

| | |
|---|---|
| TOTHHLD | –Total des ménages |
| OWNED | –Logements possédés |
| MALEHM | –Ménages avec soutien de sexe masculin |
| SINGDET | –Maisons individuelles |
| MOVABLE | –Logements mobiles |
| APT5PL | –Appartements dans un immeuble ≥ 5 étages |
| OTHDWLS | –Autres logements |
| HHSIZE1 | –Ménages d'une personne |
| HHSIZE2 | –Ménages de 2 personnes |
| HHSIZE3 | –Ménages de 3 personnes |
| HHSIZE4 | –Ménages de 4 personnes |
| HHSIZE5 | –Ménages de 5 personnes |
| HHSIZEG6 | –Ménages de 6 personnes ou plus |
| AGEHM24 | –Ménages avec soutien ≤ 24 ans |
| AGEHM34 | –Ménages avec soutien 25 à 34 ans |
| AGEHM44 | –Ménages avec soutien 35 à 44 ans |
| AGEHM54 | –Ménages avec soutien 45 à 54 ans |
| AGEHM64 | –Ménages avec soutien 55 à 64 ans |
| AGEHM74 | –Ménages avec soutien 65 à 74 ans |
| AGEHM75P | –Ménages avec soutien ≥ 75 ans |
| FAMCHLD0 | –Familles de recensement sans enfant à la maison |
| FAMCHLD1 | –Familles de recensement, un enfant à la maison |
| FAMCHLD2 | –Familles de recensement, deux enfants à la maison |
| FAMCHLD3 | –Familles de recensement, trois enfants à la maison |
| FAMCHGE4 | –Familles de recensement, quatre enfants ou plus à la maison |
| AGECLE5 | –Familles de recensement, tous les enfants à la maison ≤ 5 ans |
| AGEC614 | –Familles de recensement, tous les enfants à la maison 6 à 14 ans |
| AGEC1517 | –Familles de recensement, tous les enfants à la maison 15 à 17 ans |
| AGEC014 | –Familles de recensement, certains enfants à la maison ≤ 5 ans, autres enfants 6 à 14 ans |
| AGEC617 | –Familles de recensement, certains enfants à la maison 6 à 14 ans, autres enfants 15 à 17 ans |
| AGECLE17 | –Familles de recensement, tous les enfants à la maison ≤ 17 ans |
| AGECGE18 | –Familles de recensement, tous les enfants à la maison ≥ 18 ans |
| AGEC1718 | –Familles de recensement, certains enfants à la maison ≤ 17 ans, autres enfants ≥ 18 ans |

# Annexe B

## Caractéristiques additionnelles étudiées

Pour les examens du biais d'échantillonnage et de la cohérence des estimations–échantillon décrits aux chapitres V et VII, un ensemble de 53 caractéristiques a été utilisé. Ces 53 caractéristiques comprenaient 46 des 62 contraintes dont la liste est donnée à l'annexe A (les contraintes exclues étant TPERGE15, MALEGE15, LONEPARF et les 13 contraintes dont les codes alphanumériques commencent par FAMCH et AGEC) ainsi que les 7 caractéristiques suivantes :

- Femmes,
- Parents seuls,
- Personnes dans les familles de recensement,
- Logements loués,
- Ménages non familiaux,
- Ménages unifamiliaux,
- Soutiens du ménage de sexe féminin.

Pour l'examen des différences absolues (tableau 7 du chapitre VI), 62 caractéristiques ont été étudiées, soit les 53 caractéristiques ci–dessus, les trois contraintes TPERGE15, MALEGE15, LONEPARF de l'annexe A, ainsi que les caractéristiques suivantes :

- Personnes mariées ou séparées,
- Pères seuls,
- Femmes de 15 ans et plus,
- Personnes hors famille de 15 ans et plus,
- Enfants de 15 ans et plus,
- Ménages avec soutien > 64 ans.

Certaines des contraintes exclues de ces études n'étaient pas utilisées comme contraintes dans le recensement de 1986. C'est pourquoi on a jugé bon de les exclure, pour que les résultats soient davantage comparables entre les recensements. Certaines des caractéristiques ajoutées étaient des contraintes «indirectes» en 1991, car elles avaient une dépendance linéaire avec des caractéristiques faisant partie des contraintes de 1991. Par exemple, les Femmes présentaient une dépendance linéaire avec les Hommes et la Population totale.

# Annexe C

## Renseignements additionnels sur les statistiques utilisées dans l'examen du biais d'échantillonnage

Soit X la valeur connue d'une caractéristique 2A au niveau de la division de recensement (DR) et soit $\hat{X}^{(0)}$ l'estimateur de Horvitz-Thompson de X. $\hat{X}^{(0)}$ a été calculé en multipliant le total-échantillon non pondéré pour la caractéristique de chaque SD échantillonné par l'inverse de la fraction de sondage des ménages réalisée pour le SD, puis en faisant la somme des résultats au niveau de la DR. Les SD non échantillonnés ont été exclus de l'analyse.

L'écart-type de $\hat{X}^{(0)}$, $std(\hat{X}^{(0)}) = \sqrt{V(\hat{X}^{(0)})}$, a été calculé selon l'hypothèse que des échantillons aléatoires simples de ménages ont été tirés indépendamment dans chaque SD (en fait, des échantillons aléatoires systématiques indépendants ont été tirés). Les variances ont donc été calculées au niveau des SD, puis on en a fait la somme au niveau de la DR. Les valeurs $S^2$ de la population, qui mesurent la variance des chiffres de population, ont été utilisées dans les calculs de la variance. Voir Cochran (1977), p. 23-24 pour les formules de variance relatives aux caractéristiques des personnes et des familles, et p. 50-52 pour les formules de variance relatives aux caractéristiques des ménages et des logements.

Puisque les $\hat{X}^{(0)}$ sont des estimateurs de Horvitz-Thompson, ils sont sans biais pour X. L'échantillonnage a été fait indépendamment dans différents SD. Par conséquent, les $\hat{X}^{(0)}$ sont la somme de n variables aléatoires indépendantes, où n est le nombre de SD dans la DR. Puisque 90 % des DR comptaient plus de 25 SD (la moyenne étant de 140), n est très élevé dans la plupart des DR. Par conséquent, selon le théorème central limite, $Z^{(0)} = (\hat{X}^{(0)}-X)/std(\hat{X}^{(0)})$ devrait avoir une distribution approximativement normale (0,1) (voir Kendall et Stuart (1963), p. 193). Ce ne serait pas le cas, toutefois, si les réponses 2B présentaient un biais important, pour quelque raison que ce soit.

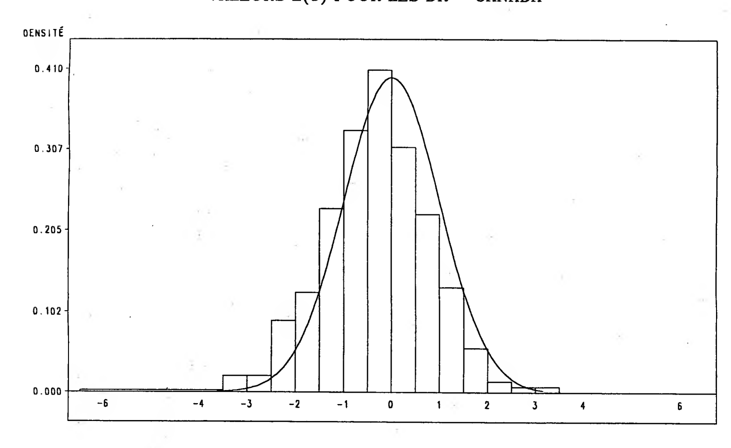
Les valeurs $Z^{(0)}$ ont été produites pour l'ensemble des 284 DR échantillonnées au Canada, à l'égard des caractéristiques 2A données au chapitre V. Afin d'évaluer la normalité des valeurs $Z^{(0)}$ au niveau des DR, des histogrammes des valeurs $Z^{(0)}$ superposés à une fonction de densité de probabilité normale ont été produits. Voir à l'annexe D des exemples de ces graphiques pour deux caractéristiques 2A.

Pour vérifier si $Z^{(0)}$ était tiré d'une distribution normale dont la moyenne est zéro (c.-à-d. si la méthode de sélection de l'échantillon était sans biais), on a calculé la moyenne $\overline{Z}^{(0)} = \sum_{i=1}^{m} Z_i^{(0)}/m$ où m = 284 (le nombre de DR) et $Z_i^{(0)}$ est la valeur de $Z^{(0)}$ pour la i[e] DR. En outre, l'écart-type des $Z_i^{(0)}$ a été déterminé, compte tenu que $std^2(Z^{(0)}) = \sum_{i=1}^{m}(Z_i^{(0)}-\overline{Z}^{(0)})/(m-1)$. Puis la statistique T donnée par $T_Z = \sqrt{m}\,\overline{Z}^{(0)}/std(Z^{(0)})$ a été calculée. Si la méthode de sélection de l'échantillon était non biaisée, T devrait suivre la distribution t de Student avec m-1 degrés de liberté. La probabilité que $|T_Z| > 1{,}960$ si la méthode de sélection de l'échantillon était non biaisée est inférieure à 0,05. Donc, si $|T_Z| > 1{,}960$, l'hypothèse que la méthode de sélection de l'échantillon était non biaisée sera rejetée et la différence entre l'estimation-échantillon et le chiffre de population sera dite statistiquement significative au niveau de 5 %.

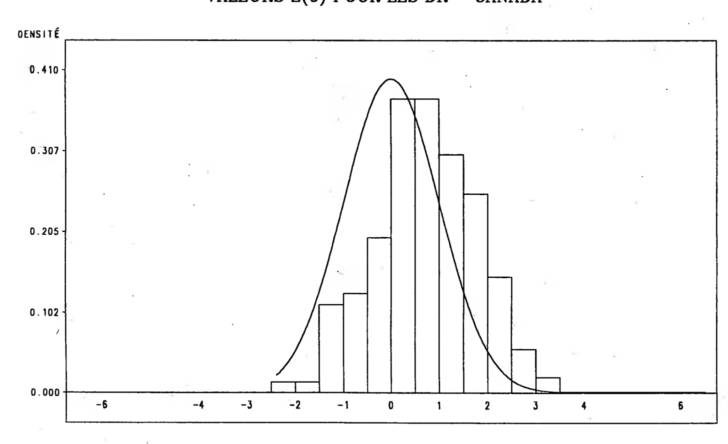
# Annexe D

## Histogrammes des valeurs $Z^{(0)}$ superposés à une fonction de densité de probabilité normale

EXAMEN DU BIAIS D'ÉCHANTILLONNAGE
SOUTIENS DU MÉNAGE DE SEXE MASCULIN
VALEURS Z(0) POUR LES DR – CANADA



EXAMEN DU BIAIS D'ÉCHANTILLONNAGE
POPULATION TOTALE
VALEURS Z(0) POUR LES DR – CANADA



# Bibliographie

Béland, Y. (1990) ***«Results and Methodology of the 1986 Sampling Variance Study»***, Statistics Canada Internal Report.

Bankier, M., Rathwell, S. and Majkowski, M. (1992) ***«Two Step Generalized Least Squares Estimation in the 1991 Canadian Census»***, Methodology Branch Working Paper, SSMD 92-007E.

Brackstone, G.J. and Rao, J.N.K. (1979) ***«An Investigation of Raking Ratio Estimators»***, Sankhya, Volume 41, Series C, Pt. 2, pp. 97-114.

Cochran, W.G. (1977) ***Sampling Techniques*** (Third Edition), New York, John Wiley & Sons.

Daoust, P. and Bankier, M. (1989) ***«Evaluation of Weighting for the 1986 Census of Population»***, Statistics Canada Internal Report.

Bureau fédéral de la statistique (1945), ***Huitième recensement du Canada, 1941, Rapport administratif du Statisticien fédéral***, Ottawa: Imprimeur du Roi.

________ (1955), ***Neuvième recensement du Canada, 1951***, vol. XI, ***Rapport administratif***, Ottawa: Imprimeur de la Reine.

________ (1968), ***Sampling in the Census***, S.M.S.03.5.

________ (1970), ***Recensement du Canada de 1961, Revue générale***, bulletin 7.2-12, nº 99-537 au catalogue, Ottawa: Imprimeur de la Reine.

Fellegi, I.P. (1964) ***«Response Variance and its Estimation»***, Journal of the American Statistical Association, 59, pp. 1016-1041.

Hansen, M.H., Hurwitz, W.N., and Bershad, M.A. (1959) ***«Measurement Errors in Censuses and Surveys»***, Bulletin of the International Statistical Institute, 38, pp. 359-374.

Kendall, M.G. and Stuart, A. (1963), ***The Advanced Theory of Statistics***, Volume 1, Charles Griffin and Company Limited, London.

Majkowski, M. (1992a) ***«1991 Census 2A/2B Discrepancies»***, Statistics Canada Internal Report.

Majkowski, M. (1992b) ***«1991 Weighting Area Formation Evaluation Report»***, Statistics Canada Internal Report.

Majkowski, M. (1994) ***«Investigation into Large Population/Estimate Differences in the 1991 Census»***, Statistics Canada Internal Report.

Royce, D. (1983) ***«The Use of Sampling in the 1981 Canadian Census»***, Statistics Canada Internal Report.

Statistique Canada (1976), ***Recensement du Canada de 1971, Revue générale***, vol. VI, partie 1, nº 99-740 au catalogue, Ottawa.

________ (1980), ***Recensement du Canada de 1976, Rapport administratif***, partie 1, nº 99-850 au catalogue, Ottawa.

________ (1980), ***Recensement du Canada de 1976, Qualité des données, Série 1: Sources d'erreurs – échantillonnage et pondération***, nº 99-844 au catalogue, Ottawa.

________ (1983), ***Recensement du Canada de 1981, Guide sommaire: Population totale***, nº 99-902 au catalogue, Ottawa.

________ (1988), ***Recensement du Canada de 1986, Recensement en bref***, nº 99-104F au catalogue, Ottawa.

________ (1990), ***Recensement du Canada de 1986, Guide à l'intention des utilisateurs sur la qualité des données du recensement de 1986 : Échantillonnage et pondération***, nº 99-136F au catalogue, Ottawa.

________ (1992), ***Recensement du Canada de 1991, Recensement en bref***, nº 99-305F au catalogue, Ottawa.

Stuart, A. (1976) ***Basic Ideas of Scientific Sampling*** (Second Edition), New York: Hafner Press.

# Centres régionaux de consultation

Les centres régionaux de consultation de Statistique Canada offrent une gamme complète des produits et services du recensement. On y trouve une bibliothèque et un comptoir de vente où les utilisateurs peuvent consulter ou acheter les différents produits, dont les publications, les disquettes de micro-ordinateurs, les microfiches et les cartes.

Le personnel des centres régionaux de consultation fournit des services de consultation et de recherche en plus du service après-vente, y compris des séminaires et des ateliers sur l'utilisation des données de Statistique Canada.

Chaque centre vous offre la possibilité d'extraire de l'information des systèmes d'extraction des données informatisés CANSIM et E-STAT de Statistique Canada. Un service de renseignements téléphonique est également disponible; ce service est gratuit pour les utilisateurs se trouvant à l'extérieur des secteurs d'appels locaux. Téléphonez, écrivez, télécopiez ou visitez le centre régional de consultation le plus près pour de plus amples renseignements.

**Région de l'Atlantique**

Dessert les provinces suivantes :
Terre-Neuve et Labrador,
Nouvelle-Écosse,
Île-du-Prince-Édouard et
Nouveau-Brunswick.

Services consultatifs
Statistique Canada
Immeuble Viking, 3e étage
Chemin Crosbie
**St. John's** (Terre-Neuve)
A1B 3P2

Appel sans frais : 1-800-565-7192
Télécopieur : (709) 772-6433

Services consultatifs
Statistique Canada
North American Life Centre
1770, rue Market
**Halifax** (Nouvelle-Écosse)
B3J 3M3

Appel sans frais : 1-800-565-7192
Appel local : (902) 426-5331
Télécopieur : (902) 426-9538

**Région du Québec**

Services consultatifs
Statistique Canada
200, boul. René-Lévesque ouest
Complexe Guy-Favreau
Bureau 412 (Tour Est)
**Montréal** (Québec)
H2Z 1X4

Appel sans frais : 1-800-361-2831
Appel local : (514) 283-5725
Télécopieur : (514) 283-9350

**Région de la Capitale nationale**

Centre de consultation statistique (RCN)
Statistique Canada
Rez-de-chaussée, immeuble R.-H.-Coats
Avenue Holland
**Ottawa** (Ontario)
K1A 0T6

Si vous demeurez à l'extérieur de la zone de communication locale, veuillez composer le numéro sans frais pour votre région.

Appel local : (613) 951-8116
Télécopieur : (613) 951-0581

**Région de l'Ontario**

Services consultatifs
Statistique Canada
Immeuble Arthur Meighen, 10e étage
25, avenue St. Clair est
**Toronto** (Ontario)
M4T 1M4

Appel sans frais : 1-800-263-1136
Appel local : (416) 973-6586
Télécopieur : (416) 973-7475

**Région du Pacifique**

Dessert la province de la Colombie-Britannique et le territoire du Yukon.

Services consultatifs
Statistique Canada
Centre Sinclair, bureau 300
757, rue Hastings ouest
**Vancouver** (Colombie-Britannique)
V6C 3C9

Appel sans frais : 1-800-663-1551
Appel local : (604) 666-3691
Télécopieur : (604) 666-4863

**Région des Prairies**

Dessert les provinces suivantes :
Manitoba, Saskatchewan, Alberta
et les Territoires du Nord-Ouest.

Services consultatifs
Statistique Canada
Immeuble MacDonald, bureau 300
344, rue Edmonton
**Winnipeg** (Manitoba)
R3B 3L9

Appel sans frais : 1-800-563-7828
Appel local : (204) 983-4020
Télécopieur : (204) 983-7543

Services consultatifs
Statistique Canada
Tour Avord, 9e étage
2002, avenue Victoria
**Regina** (Saskatchewan)
S4P 0R7

Appel sans frais : 1-800-563-7828
Appel local : (306) 780-5405
Télécopieur : (306) 780-5403

Services consultatifs
Statistique Canada
First Street Plaza, bureau 401
138 – 4e Avenue sud-est
**Calgary** (Alberta)
T2G 4Z6

Appel sans frais : 1-800-563-7828
Appel local : (403) 292-6717
Télécopieur : (403) 292-4958

Services consultatifs
Statistique Canada
Park Square, 8e étage
10001, Bellamy Hill
**Edmonton** (Alberta)
T5J 3B6

Appel sans frais : 1-800-563-7828
Appel local : (403) 495-3027
Télécopieur : (403) 495-5318

**Appareils de télécommunications pour les malentendants : 1-800-363-7629**

**Numéro sans frais pour commander seulement (Canada et États-Unis) : 1-800-267-6677**

# Bibliothèques de dépôt

La bibliothèque de Statistique Canada, située à Ottawa, tient à jour les enregistrements complets de nature courante ou historique de l'ensemble des publications de Statistique Canada, qu'elles figurent au catalogue ou non. Les employés de la bibliothèque sont à la disposition des utilisateurs pour aider ces derniers à trouver les informations dont ils ont besoin.

Bibliothèque de Statistique Canada
Immeuble R.H. Coats, 2e étage
Parc Tunney
Ottawa (Ontario)
K1A 0T6
Appel local: 613-951-8219/20
Télécopieur: 1-613-951-0939

La liste qui suit comprend les bibliothèques de dépôt complet qui reçoivent toutes les publications de Statistique Canada et de toute l'administration fédérale.

## Canada

### Terre-Neuve

**St. John's**

Memorial University of Newfoundland
Queen Elizabeth II Library
St. John's (Terre-Neuve)
A1B 3Y1

### Île-du-Prince-Édouard

**Charlottetown**

Bibliothèque des services gouvernementaux
Charlottetown (Île-du-Prince-Édouard)
C1A 3T2

### Nouvelle-Écosse

**Halifax**

Dalhousie University
Killam Memorial Library
Halifax (Nouvelle-Écosse)
B3H 4H8

**Wolfville**

Acadia University
Vaughan Memorial Library
Wolfville (Nouvelle-Écosse)
B0P 1X0

### Nouveau-Brunswick

**Fredericton**

Bibliothèque de l'Assemblée législative
Fredericton (Nouveau-Brunswick)
E3B 5H1

University of New Brunswick
Harriet Irving Library
Fredericton (Nouveau-Brunswick)
E3B 5H5

**Moncton**

Université de Moncton
Bibliothèque Champlain
Moncton (Nouveau-Brunswick)
E1A 3E9

**Sackville**

Mount Allison University
Ralph Pickard Bell Library
Sackville (Nouveau-Brunswick)
E0A 3C0

### Québec

**Montréal**

Bibliothèque municipale de Montréal
Montréal (Québec)
H2L 1L9

Services documentaires multimédia
Montréal (Québec)
H2C 1T1

Concordia University
Library
Montréal (Québec)
H3G 1M8

McGill University
McLennan Library
Montréal (Québec)
H3A 1Y1

Université de Montréal
Bibliothèque des sciences humaines et sociales
Montréal (Québec)
H3C 3T2

Université du Québec à Montréal
Bibliothèque
Montréal (Québec)
H2L 4S6

**Québec**

Bibliothèque de l'Assemblée nationale
Québec (Québec)
G1A 1A5

**Sherbrooke**

Université de Sherbrooke
Bibliothèque générale
Cité universitaire
Sherbrooke (Québec)
J1K 2R1

**Sainte-Foy**

Université Laval
Bibliothèque générale
Sainte-Foy (Québec)
G1K 7P4

### Ontario

**Downsview**

York University
Scott Library
Downsview (Ontario)
M3J 2R6

**Guelph**

University of Guelph
Library
Guelph (Ontario)
N1G 2W1

**Hamilton**

Bibliothèque publique d'Hamilton
Hamilton (Ontario)
L8R 3K1

McMaster University
Mills Memorial Library
Hamilton (Ontario)
L8S 4L6

**Kingston**

Queen's University at Kingston
Douglas Library
Kingston (Ontario)
K7L 3N6

**London**

The University of Western Ontario
D.B. Weldon Library
London (Ontario)
N6A 3K7

**Ottawa**

Bibliothèque du Parlement
Section de l'information gouvernementale canadienne
Ottawa (Ontario)
K1A 0A9

Bibliothèque nationale du Canada
Ottawa (Ontario)
K1A 0N4

Université d'Ottawa
Bibliothèque Morisset
Ottawa (Ontario)
K1N 9A5

**Sudbury**

Université Laurentienne de Sudbury
Bibliothèque
Sudbury (Ontario)
P3C 2C6

**Thunder Bay**

Lakehead University
Chancellor Paterson Library
Thunder Bay (Ontario)
P7B 5E1

Bibliothèque publique de Thunder Bay
Thunder Bay (Ontario)
P7E 1C2

**Toronto**

Bibliothèque de l'Assemblée législative
Toronto (Ontario)
M5S 1A5

Metropolitan Toronto Reference Library
Toronto (Ontario)
M4W 2G8

University of Toronto
Robarts Library
Toronto (Ontario)
M5S 1A5

**Waterloo**

University of Waterloo
Dana Porter Arts Library
Waterloo (Ontario)
N2L 3G1

**Windsor**

Bibliothèque publique de Windsor
Windsor (Ontario)
N9A 4M9

## Manitoba

**Winnipeg**

Bibliothèque de l'Assemblée législative
Winnipeg (Manitoba)
R3C 0V8

The University of Manitoba
Elizabeth Dafoe Library
Winnipeg (Manitoba)
R3T 2N2

## Saskatchewan

**Regina**

Bibliothèque de l'Assemblée législative
Regina (Saskatchewan)
S4S 0B3

**Saskatoon**

University of Saskatchewan
The Main Library
Saskatoon (Saskatchewan)
S7N 0W0

## Alberta

**Calgary**

The University of Calgary
MacKimmie Library
Calgary (Alberta)
T2N 1N4

**Edmonton**

Bibliothèque publique d'Edmonton
Edmonton (Alberta)
T5J 2V4

Bibliothèque de l'Assemblée législative
Edmonton (Alberta)
T5K 2B6

The University of Alberta
Library
Edmonton (Alberta)
T6G 2J8

## Colombie-Britannique

**Burnaby**

Simon Fraser University
Library
Burnaby (Colombie-Britannique)
V5A 1S6

**Vancouver**

The University of British Columbia
Library
Vancouver (Colombie-Britannique)
V6T 1Y3

Bibliothèque publique de Vancouver
Vancouver (Colombie-Britannique)
V6Z 1X5

**Victoria**

Bibliothèque de l'Assemblée législative
Victoria (Colombie-Britannique)
V8V 1X4

University of Victoria
McPherson Library
Victoria (Colombie-Britannique)
V8W 3H5

## Territoires du Nord-Ouest

**Yellowknife**

Bibliothèque gouvernementale des Territoires du Nord-Ouest
Yellowknife (Territoires du Nord-Ouest)
X0E 1H0

## Autres pays

## République fédérale d'Allemagne

Preussischer Kulturbesitz
Staatsbibliothek
Abt. Amtsdruckchriften U. Tausch
Postfach 1407
1000 Berlin 30
Allemagne

## Royaume-Uni

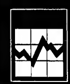
The British Library
Londres, WC1B 3DG
Angleterre, Royaume-Uni

## Japon

National Diet Library
Tokyo, Japon

## États-Unis d'Amérique

Library of Congress
Washington, D.C. 20540
États-Unis d'Amérique



# BON DE COMMANDE

Statistique Canada

| ENVOYEZ À : | COMPOSEZ : | TÉLÉCOPIEZ AU : |
|---|---|---|
| **Division du marketing**<br>**Vente des publications**<br>**Statistique Canada**<br>**Ottawa (Ontario)**<br>**Canada K1A 0T6** | **1-800-267-6677**<br>Faites débiter votre compte VISA ou MasterCard. De l'extérieur du Canada et des États-Unis, composez le (613) 951-7277. Veuillez ne pas envoyer de confirmation. | **(613) 951-1584**<br>VISA, MasterCard et bon de commande seulement. Veuillez ne pas envoyer de confirmation; le bon télécopié tient lieu de commande originale. |

(*Veuillez écrire en caractères d'imprimerie.*)

Compagnie

Service

À l'attention de     Fonction

Adresse

Ville     Province

Code postal     Téléphone ( )     Télécopieur ( )

Veuillez vous assurer de remplir le bon **au complet**.

**MODALITÉS DE PAIEMENT :**

(*Cochez une seule case*)

☐ **Veuillez débiter mon compte** ☐ VISA ☐ MasterCard

N° de carte

Signature     Date d'expiration

☐ **Paiement inclus** ______ $

Veuillez faire parvenir votre chèque ou mandat-poste à l'ordre du *Receveur général du Canada – Publications*.

☐ **N° du bon de commande**

*(Veuillez joindre le bon)*

Signature de la personne autorisée

| Numéro au catalogue | Titre | Édition demandée ou Inscrire "A" pour les abonnements | Abonnement annuel ou prix de la publication | | | Quantité | Total $ |
|---|---|---|---|---|---|---|---|
| | | | Canada $ | États-Unis $ US | Autres pays $ US | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

| | |
|---|---|
| TOTAL | |
| RÉDUCTION (s'il y a lieu) | |
| TPS (7 %) (Clients canadiens seulement) | |
| TOTAL GÉNÉRAL | |

▶ **Veuillez noter que les prix au catalogue pour les É.-U. et les autres pays sont donnés en dollars américains.**

▶ TPS N° R121491807

▶ Le chèque ou mandat-poste doit être établi à l'ordre du *Receveur général du Canada – Publications*.

▶ Les clients canadiens paient en dollars canadiens et ajoutent la TPS de 7 %. Les clients à l'étranger paient le montant total en dollars US tirés sur une banque américaine.

**PF 093238**






Statistique Canada     Statistics Canada

